Psychometrika

A JOURNAL DEVOTED TO THE DEVELOPMENT OF PSYCHOLOGY AS A QUANTITATIVE RATIONAL SCIENCE

THE PSYCHOMETRIC SOCIETY · ORGANIZED IN 1935

VOLUME 16
NUMBER 4
DECEMBER
1 9 5 1

PSYCHOMETRIKA, the official journal of the Psychometric Society, is devoted to the development of psychology as a quantitative rational science. Issued four times a year, on March 15, June 15, September 15, and December 15.

DECEMBER 1951 VOLUME 16, NO. 4

Printed for the Psychometric Society at 23 West Colorado Avenue, Colorado Springs, Colorado. Entered as second class matter, September 17, 1940, at the Post Office of Colorado Springs, Colorado, under the act of March 3, 1879. Editorial Office, Department of Psychology, The University of North Carolina, Chapel Hill, North Carolina.

Subscription Price: The regular subscription rate is $10.00 per volume. The subscriber receives each issue as it comes out, and a second complete set for binding at the end of the year. All annual subscriptions start with the March issue and cover the calendar year. All back issues are available. The price is $1.25 per issue or $5.00 per volume (one set only). Members of the Psychometric Society pay annual dues of $5.00, of which $4.50 is in payment of a subscription to *Psychometrika*. Student members of the Psychometric Society pay annual dues of $3.00, of which $2.70 is in payment for the journal.

Application for membership and student membership in the Psychometric Society, together with a check for dues for the calendar year in which application is made, should be sent to

RAYMOND A. KATZELL, Chairman of the Membership Committee
Psychological Services Center, Syracuse University, Syracuse 10, New York

Payments: All bills and orders are payable in advance. Checks covering membership dues should be made payable to the *Psychometric Society*. Checks covering regular subscription to *Psychometrika* and back issue orders should be made payable to the *Psychometric Corporation*. All checks, notices of change of address, and business communications should be addressed to

ROBERT L. THORNDIKE, Treasurer, Psychometric Society and Psychometric Corporation
Teachers College, Columbia University
New York 27, New York

Articles on the following subjects are published in *Psychometrika*:

(1) the development of quantitative rationale for the solution of psychological problems;
(2) general theoretical articles on quantitative methodology in the social and biological sciences;
(3) new mathematical and statistical techniques for the evaluation of psychological data;
(4) aids in the application of statistical techniques, such as nomographs, tables, work-sheet layouts, forms, and apparatus;
(5) critiques or reviews of significant studies involving the use of quantitative techniques.

The emphasis is to be placed on articles of type (1), in so far as articles of this type are available.

In the selection of the articles to be printed in *Psychometrika*, an effort is made to obtain objectivity of choice. All manuscripts are received by one person, who

(*Continued on the back inside cover page*)

Psychometrika

CONTENTS

VOLUME SIXTEEN DECEMBER 1951 NUMBER FOUR

ELECTION OF PRESIDENT OF THE PSYCHOMETRIC SOCIETY

The Committee on Elections of the Psychometric Society announces that John C. Flanagan has been elected President of the Psychometric Society for the year 1951-192.

PSYCHOMETRIC MONOGRAPHS

A new issue of the Psychometric Monograph Series, No. 5, *The Description of Aptitude and Achievement Tests in Terms of Rotated Factors*, by John W. French, is now available at $4.00. Your order for it (or for any previous issue) should be sent to

The University of Chicago Press
5750 Ellis Avenue
Chicago, Illinois

Earlier issues of the Psychometric Monograph Series, which are still available, are as follows:

Thurstone, L. L. Primary mental abilities. *Psychometric Monograph No. 1*, $2.00.

Thurstone, L. L. and Thurstone, Thelma Gwinn. Factorial studies of intelligence. *Psychometric Monograph No. 2*, $1.50.

Wolfle, Dael. Factor analysis to 1940. *Psychometric Monograph No. 3*, $1.25.
Thurstone, L. L. A factorial study of perception. *Psychometric Monograph No. 4*, $2.50.

The Psychometric Monograph Committee is composed of J. P. Guilford, *Chairman*; L. L. Thurstone, Harold Gulliksen, Paul Hort, and Frederic Kuder. Manuscripts and correspondence for this series should be addressed to:

J. P. GUILFORD, *Chairman*
Psychometric Monograph Committee
Box 1134
Beverly Hills, California

PSYCHOMETRIKA—VOL. 16, NO. 4
DECEMBER, 1951

THE FACTORS IN FACTORING BEHAVIOR*

QUINN MCNEMAR
STANFORD UNIVERSITY

Twenty years have now elapsed since Professor Thurstone's ingenuity pulled the factor problem out of its tetrad difference quagmire. Most of us have watched Thurstone's brain-child grow. One might say that the fledgling was so precocious as to reach maturity during the first ten years of its life. Its development was fostered not only by Thurstone but also by the resistance it encountered. Indeed, even as an infant, it was forced to throw aside its swaddling clothes in order to kick back at the Spearmans, the Tryons, the Anastasis, and others who sometimes rightly, ofttimes wrongly, pointed a finger at supposed imperfections. The youngster weathered more or less successfully the many storms it had stirred up, and by the end of its tenth year of life had in the hands of Thurstone made significant contributions and was ready for the sober coming-of-age evaluation of its strengths and weaknesses by Dael Wolfle. We suspect that the facts of life pointed out by Wolfle had already been whispered to the growing adolescent by none other than Thurstone himself.

At any rate we shall presume that the new method had reached scientific maturity by its tenth birthday. What about the scientific maturity of the users of the new technique? This question is prompted by the apparent fact that after twenty years of factoring there is altogether too little acceptance of the method and the results obtained thereby.

It occurred to the speaker that those who reject factor analysis may be doing so because of the manner of its use; therefore as a first step we determined to make a study of the factors in factoring behavior. Are there factors in the behavior of the factorists that might mitigate against their acceptance by the non-factorists? This is really a question in social psychology: Why hasn't the minority group known as factorists been accepted by the majority? The social psychologist usually seeks an answer in the characteristics of the majority, but we propose to study the minority group.

*Address of the Retiring President of the Psychometric Society, delivered at Chicago, September 1, 1951.

Our starting matrix contains observations on behavior as culled from 73 reports written by 70 different authors during the ten year period, 1941-1950. We are convinced that our sample is representative of the universe. In fact, it is nearly the entire universe of the papers appearing during the decade. We start with no hypotheses as to the factors in the domain of factoring behavior. We merely assume that the factor method can be used to study the factorists.

We should not have to tell such a sophisticated audience as this that the wealth of material in our starting behavioral matrix will permit the application of either the R-technique or the Q-technique. Indeed, there is enough recorded behavior on certain persons to permit the use of the P-technique. The Q-technique, however, seems more appropriate for our purpose. Time will not permit us to tell you about the rotations, gyrations, inversions, projections, and reflections used in isolating a factor. It is not my fault that at times I had to stand on my head in order to analyze some of the upside down behavior which I found.

Before proceeding to our results we should insert a word concerning the subjects who served in this experiment. We can't say that all were willing volunteers — some may have been pressured into publication. At risk of being called an ungrateful scoundrel, I must admit that I am not thankful for the voluminous amount of behavior supplied by certain of the subjects. Finally, I must declare that I do not wish to reveal the identity of any of my cases. If a bit of behavior described herein resembles that of any of you or of your friends, please remember that none of the behavior mentioned bears any resemblance to that of persons, living or dead, who lived prior to Galton.

Now to the factors in factoring behavior. The first problem we face is that of symbols for the various factors. Surely the idea of factor analysis must have occurred to the Greeks, hence we will use the Greek alphabet and thus add to the scientific respectability of the several factors which we are about to describe. All told, ten major factors were found, plus minor ones which we relegate to a residual plane.

Factor Alpha involves a cluster of 13 persons who seem alike in that they enjoy the bouncy life — bouncing around because their starting matrices are based on samples too small to permit stability. For these 13 individuals, all using the R-technique, the average loading on this factor is 56; that is, their average sample size is 56. Those of our colleagues who know something of the sampling instability of correlation coefficients cannot be expected to place much credence in a method which they find being used on samples as small as 30.

Factor Beta is oblique to factor Alpha, but it is more difficult to identify. The people in this cluster seem to have in common a lack

of appreciation of certain sampling aspects of the factorial method, but the manifest behavior varies somewhat from person to person. Some of our factorists eliminate a variable which shows near zero correlations with the other variables, some retain such a variable. When interpreting factors some of our subjects use tests with loadings down to .40, others go down to .30, others to .20, and some to intermediate values, but these lower bounds bear no relationship whatsoever to the sample size. One of our cases got into the Beta cluster on the basis of unique behavior — for a sample of 100 he kept extracting factors until the median residual was down to .013, a low in factoring behavior.

Factor Gamma has to do with rotational behavior, but the factor is bipolar — you either rotate or you don't rotate. Careful examination of the records reveals that the decision to or not to rotate is not always a rational one. There are those who will allow the first centroid to stand as a general factor despite an abundance of zero intercorrelations, and there are those who insist on rotating — they seem to loathe a general factor even in situations for which it seems sensible to postulate such a factor. Four of our subjects proclaim emphatically that for their variables there is no general factor even though their correlations are all appreciably positive. Such dogmatic statements can only be matched by exhuming the ghost of Spearman. It is possible to argue that our rotational factor is tripolar if we bring into the picture those who set up oblique axes, then proceed to the extraction of second-order factors and the grudging admission that there might after all be such a thing as a general factor.

The fourth factor, Delta, is readily identifiable and occurs with sufficient frequency to provide nonfactorists with ample reason for being skeptical of factorists. Let's look at a few illustrations so that you can judge for yourself whether such behavior adds to the prestige of factor analysis. Case A uses a battery with a median reliability of .52; Case B uses variables with reliabilities as low as .43, .41, .40, and .29; Case C does not omit a test with a reliability of .21; Case D includes a test with a reliability of .18; and Case E uses six tests with reliabilities below .30. Now hold on to your seats as we zip to a new low — on one of Case E's tests had a reliability of .08. We wouldn't call this ridiculous. Unfortunately, our experimental friends are not sophisticated enough to know that error variance simply emerges from the factor mill as error variance.

In mathematical texts, Epsilon is apt to represent a very small quantity, but our Epsilon factor involves a sizable number of individuals who seem to have forgotten that there are a few restrictive assumptions in the fundamental factor equation. Twelve of our cases have *hung* themselves in the Epsilon cluster by ignoring the fact that

the mathematical model calls for experimental independence among the variables to be analyzed. Such lack of experimental independence can arise by way of the halo effect in ratings, or when part-whole correlations are involved, or when correlations between variables and various ratios of the variables are included. The most readily visible consequence of analyses based on various types of spurious correlations is the emergence of communalities in excess of unity. This happens for 12 of our subjects five of whom present us with communalities in excess of 1.2, the highest being 1.56. No amount of revising of starting estimates will ever reduce these to acceptable values. The presence of spurious correlations can also affect the factorial space. Rorschach variables illustrate the point. Is it surprising that a Rorschach variable defined as the ratio of W to M and one defined as the ratio of M to C should come out with high opposite-signed loadings on the same factor?

Factor Zeta had us puzzled for a long time despite the fact that 11 cases clearly belonged in this cluster. It seemed that we should be able, by diligent search of the dictionary, to find an appropriate name for the observed common behavior of those in this cluster. We searched in vain. In desperation we sought a clinical psychologist, who scornfully suggested that we make case studies of the 11 subjects to find out what they had in common other than factoring behavior. The case records revealed that all were Ph. D.'s, but that fact didn't seem to help a bit. The only other common characteristic in their life histories was that all were Presbyterians. Just as I was about to turn to dianetics I had a flash of Hubbardian insight: Predestination. All had obviously predestined the results of their factor analyses. To illustrate: in go four achievement tests in psychology, out comes an achievement factor. In go three measures of attitude X, out jumps attitude X as a factor. In go four addition tests, and an addition factor emerges. In go nine tests of nine defined skills and, by judicious choice of the principal component method, nine factors are found, as was postulated. In go four body sway tests, and behold a body sway factor. A battery consists of 2 memory tests, 3 perceptual tests, 6 attention tests, 2 space tests, and 2 number tests. A factor for each type of test comes out. And so on. Is it any wonder that some of our non-factorists are willing to believe the old saw that "one gets out of factor analysis what one puts in"?

Closely related to predestination is our next factor, Eta. It consists of a small cluster of persons who factored domains too limited to demonstrate anything, and a cluster of five, each of whose domains was so extensive as to include areas know to be uncorrelated. Do we need a factor analysis for a battery consisting of six measures of general intelligence, five mechanical tests, and six Seashore music

subtests? What is to be gained by dumping into the mill both interest tests and measures of personality of the MMPI variety? What do we expect to get when physiological measures, personality tests, attitude scales, and intellectual tests are tossed in together?

Our eighth factor, Theta, needs to be investigated further by the P-technique. The individuals involved show a peculiar and difficult-to-understand type of self-inconsistent behavior when they choose variables for identifying and interpreting their factors. Supposedly this should be done on the basis of the magnitude of factor loadings, but six of our subjects seem to make their own unstated rules as they go. One sets a loading of .25 as a limit, goes as low as .20 at times and at other times ignores tests with loadings of .25, .26, .28, .29, and .31. Another goes down to a .23, overlooks two other .23's, two .24's, and a .30. Even for the same factor a .23 is included but a .24 is ignored. In a third instance we find that, for the interpretation of one factor, a test with a loading of .29 is used, and one with a loading of .32 is not used. The fourth subject sets .30 as his limit, yet goes down to .22 when convenient. Our fifth case goes down to .20, while ignoring loadings of .27 and .31. The last in this cluster is both quantitatively and qualitatively different from the others — he goes down to .05 and offers an explanation for the test having that much of a loading. We wonder whether Professor Thurstone ever dared dream that his method would be used by a person with such super-insight as to explain why a factor accounts for one fourth of one per cent of the variance of a test. Can we hope that such precision of thought will convince the skeptic that factor analysis is a precise method? One could dismiss this rare bit of behavior as that of an untrained novice if it were not for the fact that he was trained under one of the most prominent factorists. Similar high-order insight is shown by some of our subjects when they succeed in defining a factor on the basis of tests which intercorrelate as low as .17. Perhaps this is one way of taking the ignorance out of a correlation symbol.

We come next to a *general* factor in factoring behavior. Despite our great effort to avoid such a factor by rotating axes, we never quite succeeded in getting rid of it. There can be no doubt about its existence — we strongly suspect that it is a factor that characterizes all factor analysts, that super-insight is not needed to detect it, and that saturations on this factor vary greatly from factorist to factorist. Of the 70 cases in our sample, 13 had very high loadings on this factor. Consideration of these cases compels one to call this general factor the "struggle" factor. No Greek letter is needed for such an obvious factor — when interpreting factors all factorists struggle and struggle and struggle.

Now, to the tenth and last factor in our list. At first we were

apprehensive lest this also be a general factor, since the finding of two first-order general factors in the same domain might well lead some to suspect our analysis. We spent many hours in an effort either to rotate it out or to break it down into subfactors. We succeeded in doing the latter, so what at first appeared to be a general factor of *hypothesizing* behavior has now been analyzed into four components; that is, with regard to the use of hypotheses our cases seem to fall into four clusters. First, there are those who apparently started with no hypotheses, used none in the process, and ended up with none. Second, there are those who started with no hypotheses, but ended with hypotheses. Third, there are those who started with hypotheses and ended believing they had proven them. These have, of course, very high loadings on the predestination factor. Fourth, there are those who start with hypotheses, use them enroute, and end with hypotheses. It is in connection with our tenth factor that the mathematical statistician raises the nasty question as to how hypotheses involving statistical material can ever be subjected to critical test in the absence of knowledge concerning sampling errors. What the mathematical statistician fails to realize is that the factorists have developed an immunity to sampling problems.

Be that as it ought to be, we must return to the behavioral realm in order to report in some detail the hypothesis behavior of a couple of our subjects. Our first case illustrates what happens when a person starts with hypotheses. For a given domain this case hypothesized that there would be four leading factors — he found seven. He set up 17 minor hypotheses and found support for only three of them. You are correct if you guessed that this person was low on the predestination factor. As a second illustration we choose a case who did not start with hypotheses — a table of intercorrelations just happened to be available. He ended up with no less than 58 hypotheses, which we presume must be a world record of some sort. Further research is needed to learn whether this subject is suffering from too much of Thurstone's fluency factor or too much of Spearman's perseveration or from simple flight of ideas. This case does prove one thing: It is still possible to publish unbridled speculations in our journals.

Summarizing briefly, the factors in factoring behavior have to do with nabbing a small sample, ignoring other crucial sampling matters, treating the rotational problem irrationally, using tests of known unreliability, violating the requirement of experimentally independent measurements, predestinating the outcome, tossing in too much or not enough, choosing and ignoring tests when naming factors, struggling to make sense out of the results, and varying all over the map in the use of hypotheses.

There is, obviously, an element of predestination in our results

in that we did not include the domain of commendable behavior for analysis. Our aim was to find at least a partial answer to the question as to why the factorial method has not gained wider acceptance. Consequently we turned to the recorded behavior of the factorists with the specific question: What does a given record contain which might lead a reader to say, "If this represents the standards among factor analysts, I'll have none of it."

The only possible conclusion is that during the second decade of mutiple-factor analysis there has been an appalling amount of factoring behavior which is not only not conducive to the acceptance of factor analysis but which also tends to bring the method into disrepute. It is obvious that the actual surface behavior of factorists described herein does not have as a source the high purposes set forth at the founding of the Psychometric Society. We as members of the Society, which includes practically all of those in America who teach quantitative methods in psychology, regard ourselves as the true proponents of the development of psychology as a quantitative rational science. It is our privilege and obligation to set and maintain high standards for research in the quantitative area. If we don't do it, who will?

PSYCHOMETRIKA—VOL. 16, NO. 4
DECEMBER, 1951

A FACTORIAL STUDY OF THE REASONING AND CLOSURE FACTORS*

WILLIAM A. BOTZUM
UNIVERSITY OF NOTRE DAME

A battery of 46 tests was given to 237 college men. A factor analysis using the Thurstone technique revealed eight clearly interpretable first-order factors, one dubious factor, and a residual factor. The factors were interpreted as induction, deduction, flexibility of closure, speed of closure, space, verbal comprehension, word fluency, and number. Four second-order factors were abstracted from the matrix of first-order correlations. The presence of induction, deduction, and flexibility of closure on the first second-order factor, interpreted as an analytic factor, confirmed previous indications of relationships between the reasoning and closure factors. A second bipolar factor is interpreted as a speed of association factor. The third factor is interpreted as facility in handling meaningful verbal materials—perhaps an ability to do abstract thinking. The fourth factor is possibly a second-order closure factor—perhaps an ability to do concrete thinking.

This study is an investigation of the relationships between the reasoning and closure factors. Since the reasoning factors have been considered cardinal elements in intelligence, their association with the closure factors, indicated in previous research (4, 10, 14), assumes considerable importance and interest. Up to the present time no thoroughgoing analysis of these relationships has been made.

In his pioneer study of the "primary mental abilities" (6) Thurstone described three reasoning factors: induction or the ability to discover an underlying rule or principle in a task, deduction or the ability to proceed logically and to apply principles, and restricted thinking or the ability to solve tasks that "involve some form of restriction in the solution." Of these three factors only one, induction, has proved itself a consistent factor in subsequent studies (7, 8, 13). A factor termed deduction was isolated in one study (8), but it probably can not be identified with the previous deduction factor. Re-

*The author is grateful to Professor L. L. Thurstone for his encouragement and invaluable advice and for permission to use many tests originally prepared in the Psychometric Laboratory of the University of Chicago, to Mr. James Degan for assistance in rotations, and to the Social Science Research Committee of the University of Chicago for a grant to this study.

stricted thinking has not appeared since the first study. Tests purporting to measure these three factors were included in the present study. The induction factor was most heavily represented.

The closure factors were first reported in the *Factorial Study of Perception* (10). There Thurstone described three closure or gestalt factors: strength of configuration, flexibility of closure, and speed of closure. Bechtoldt in a subsequent study of the perceptual domain (1) found evidence of two closure factors, one a "facility in restructuring formal perceptual material possessing a weak intrinsic structure," and the other, a "facility in organizing simultaneous visual configurations under the distraction of continuing activity." Other investigators have reported factors similar to the above closure factors (3, 4, 14). As will be shown in the discussion of the factorial results of the present study, two of the closure factors, flexibility of closure and speed of closure, seem well established. Their role in the complex of mental organizations is just beginning to be delineated.

In the battery of the *Factorial Study of Perception* (10) Thurstone included composite tests of each of the recognized primary mental abilities. The composite test of induction had prominent loadings on the flexibility of closure factor, which Thurstone described as "the ability to manipulate several more or less irrelevant or conflicting gestalts or configurations." Since this was the only test of reasoning in the battery, no reasoning factor appeared in the study and the indication of a relationship between the reasoning and the closure factors was suggestive rather than conclusive. The present study employs probably the largest group of reasoning tests ever to be assembled in a single battery together with representative tests of the closure factors. All three of the reasoning factors and the five closure factors were represented by tests. In addition, to help stabilize the reasoning factors, which are invariably complex tests involving other factors than reasoning, tests of four of the stable primary mental abilities, space, number, verbal comprehension, and verbal fluency, were added to the battery.

The Tests

There are forty-six tests in the present battery.* All of them are group tests of the paper and pencil variety, speed tests rather than power tests. Preceding each test proper, with one or two exceptions, is a fore-test which familiarizes the subject with the task demanded.

*A microfilm copy of these tests may be secured from the department of microfilming at the University of Chicago.

The completion type of test item form was employed throughout the test battery wherever feasible and especially in important tests for key factors, because it was felt that better than the multiple-choice test item, the completion form approaches the actual situation in which the abilities in question are called into play.

Most of the tests were reproductions or modifications of tests utilized in previous studies, and full accounts of them can be obtained in the references quoted following each test. Moreover in the discussion of the factorial results many of the tests will be described in some detail.

Twenty-two of the forty-six tests are reasoning tests: Letter Series (8), Number Series (6), Letter Grouping (8), Number Patterns (8), Reasoning II (adapted from Cyril Burt for the Hyde Park study of induction, (8), Pattern Analogies (6), Reasoning III (adapted from Cyril Burt for the Hyde Park study of induction (8), Secret Writing (13), Arithmetic (6), Tabular Squares (a new test involving the filling in of tables of numbers), Tabular Completion (6), Marks (8), Numerical Judgment (6), False Premises (6), Figure Classification (6), Reasoning I (6), Verbal Analogies I (6), Verbal Analogies II (a new test that requires the selection of a complete ratio rather than the second half of a ratio), Figure Grouping (8).

There are fourteen tests of closure in the battery: Copying (6), Gottschaldt Figures (10), Designs (7), Block Counting (6), Paper Puzzles (prepared by T. G. Thurstone, and similar to the form board test used in 6), Mechanical Movements (6), Hidden Words (a new test in which the subject finds four letter words amidst a jumble of letters), Street Gestalt (10), Backward Writing (13), Mutilated Words (10), Incomplete Words (13), Four-Letter Words (1), Scrambled Words (a new test in which the subject identifies four-letter words, the letters of which have been rearranged to form a meaningless but pronounceable word), Hidden Letters (10), Picture Squares (prepared by T. G. Thurstone and used in 1), Hidden Pictures (10), Identical Forms (6).

The following tests were added to the battery to anchor the various reference factors: (a) tests of space: Figures (13), Cards (13), Solid Blocks; (b) tests of verbal comprehension: Definition (a variation of the Completion test of the American Council on Education Psychological Examination), Vocabulary (6), Completion (a sentence completion type of test constructed for the present battery); (c) tests of word fluency: First Letter (13), Suffixes (13); (d) tests of number: Multiplication (6), Addition (6).

Testing and First-Order Factoring

Some of the tests were administered in a preliminary tryout. Then two hundred and fifty students at the University of Notre Dame volunteered to participate in the study proper, and of this number two hundred and thirty-seven finished all forty-six tests. The scores of these latter were reduced to single digits, and the product moment correlation coefficients and the split-half reliability coefficients contained in Table 1 were computed with the assistance of International Business Machine equipment and a Marchant calculator. The range of the correlation coefficients is from —.11 to .76, and eighty-three per cent of all the coefficients are significant at the 1% level of confidence. It should be noted that none of the few negative coefficients is significant. In the factor analysis the correlation coefficients were carried out to four places, but they have been reduced to two digits and the decimal points omitted in Table 1 because of the size of the table.

The split-half method of ascertaining the reliability of speeded tests yields coefficients that are admittedly higher than those obtained when the scores from equivalent forms of the tests are correlated. In the absence of such equivalent forms for the forty-six tests of the present experimental battery, the split-half coefficients, which are generally quite high, are presented as indicative of the probable range of the corresponding true reliability coefficients. Since a factor analysis employs the inter-test correlations and not the reliability coefficients, the latter are not of primary importance in the present study. If tests are sufficiently reliable that meaningful factors can be obtained—as seems to be the case in hand—one can later determine the reliability of the tests by more acceptable procedures. These split-half reliability coefficients are, therefore, reported for what they are worth with the admission that some of them are probably too high.

The correlation matrix was factored twice by the complete centroid method of Thurstone (12), so as to stabilize the communalities. Ten factors were abstracted leaving negligible residuals. The unrotated factor matrix is reproduced in Table 2. This factor matrix was rotated to a first-order oblique simple structure according to the principles and methods of rotation set down by Thurstone (12). The final transformation matrix which carried the orthogonal F matrix into the oblique solution is found in Table 3. The resulting oblique factor matrix is reproduced in Table 5, and the cosines between the reference vectors or axes of this oblique factor matrix are given in Table 4.

The Interpretation of the Factors

In the case of each factor, all tests having loadings of more than .20 will be listed in the table preceding the interpretation of the factor. The interpretation of each factor is based primarily upon the tests with high loadings, i.e., loadings above .30. When a structure is highly oblique, however, as in the present study, one may not expect the factor loadings to be as high as in an orthogonal pattern. Hence greater liberty may be taken in interpreting factor loadings of lesser magnitude. Such interpretations of loadings between .20 and .30 will be offered as supplementary and subordinate evidence, when the interpretations seem reasonable, and particularly when the interpretations are in accord with results reported in other studies employing the same or very similar tests.

Factor A: Induction

1.	Letter Series	.47
2.	Number Series	.45
3.	Letter Grouping	.38
4.	Number Patterns	.36
5.	Reasoning II	.35
6.	Pattern Analogies	.33
7.	Reasoning III	.33
8.	Secret Writing	.31
9.	Arithmetic	.31
10.	Tabular Squares	.29
11.	Tabular Completion	.27
12.	Marks	.24
13.	Numerical Judgment	.23
18.	Verbal Analogies II	.22
26.	Hidden Words	.21

Introspection of the processes involved in these tests clearly indicates induction as their chief component. In Letter Series one must analyze the arrangement of the letters to determine the underlying principle of construction and then fill in the empty blanks with the appropriate letters. The same process is involved in Number Series. In Letter Grouping, the subject must find something common to three of the four groups, a principle of grouping. In Number Patterns, the digits in each cell of the square are selected according to some principle of arrangement, which the subject must discover before he can fill in the empty cell designated by an 'x.' Again induction. In each of the problems in Reasoning II and Reasoning III, the subject must find the principle or reason which give the key to the solution. In Pattern Analogies, the subject must determine the principle or rela-

tionship between the first two members of the ratio before he can complete the second pair.

In the remainder of the tests having smaller loadings on this factor, it is more difficult to single out the inductive element. In Secret Writing, the subject has to break the code for each of the items. A test of this type did have a loading on the induction factor in the Thurstones' *Factorial Studies of Intelligence* (13). Arithmetic probably involves the induction of a principle for each of the problems in the test. In the primary mental abilities study, henceforth referred to as the PMA study (6), this test had a loading on the induction factor. The most significant loading of Tabular Completion in the PMA study (6) was on the inductive factor, and hence it is not surprising to find it with a loading on factor *A* in this battery. Since Tabular Squares is very similar to it, we should expect it also to have a loading on induction. In each of the problems in this test, the subject must find the key or principle that begins the solution of the whole item. The quicker this is found the easier the solution. In Marks, the principle of marking the first five lines must be found in order to mark the sixth line of the item. In the PMA study (6) Numerical Judgment had a loading on the induction factor. Apparently the subject looks for a short cut for estimating an answer, and this amounts to finding a principle for each of the items. Verbal Analogies II like Pattern Analogies demands the discovery of a relationship between the members of the first ratio. Unfortunately, the verbal element so dominates in this test that with college students the inductive element is largely submerged. Hidden Words has only a small loading on the inductive factor. It would seem that the person who finds some principle to help him discover the Hidden Words does the task best. Such a principle might be the detection of straight lines of letters.

Factor *A* then is clearly induction. Regarding this factor two observations are in order. First of all it will be noted that the factor is a function not limited to a particular type of material but transcends the material. Number Series, Arithmetic, Tabular Squares, Tabular Completion, Number Patterns, Numerical Judgment, deal with numbers. Letter Series and Letter Grouping work with letters. Reasoning II and Reasoning III, Verbal Analogies II, and Arithmetic to a large extent, are verbal tests. Hidden Words and Secret Writing may be said to involve words to a lesser degree. Pattern Analogies involves forms and figures. The second observation follows logically upon the first: the loadings on the inductive factor are not high. In rotating to an oblique solution tests that are complex tend to have their loadings on the several factors depressed. The ideal at which

one aims in the construction of tests is the so-called "pure" test, a test whose variance is almost entirely explained by one factor. Unfortunately it seems that pure tests of induction cannot be constructed, since by its nature a test of induction demands a medium of operation, which necessarily brings into play other factors.

One might have expected Figure Grouping to be represented on Induction. This variant of Figure Classification, however, has never been a satisfying test. Constructed for the Hyde Park Study (8), it gave only a moderate loading on induction. In the Thurstone's *Factorial Studies of Intelligence* (13), it had no loading on induction, and only a small loading on perception. In the recent study of mechanical aptitude (12), it showed no significant loadings on any factor. And yet its communality is about .65. Apparently this test is one of those mentioned above, whose factor loadings are depressed in an oblique solution.

Mention may be made here of the Army Air Force study of the reasoning factors (2). Three possible reasoning factors were reported in this study, a general reasoning factor, and two additional factors whose interpretation was merely conjectural. There is no apparent correspondence between these factors and the reasoning factors reaffirmed in the present study. Perhaps an oblique rather than an orthogonal simple structure solution of the Army Air Force data would yield more similar results. In any case the interpretation of the present reasoning factors is very plausible and agrees with previous studies employing the same or very similar tests.

Factor B: Deduction

14.	False Premises	.42
15.	Figure Classification	.42
16.	Reasoning I	.40
17.	Verbal Analogies I	.26
18.	Verbal Analogies II	.24
19.	Figure Grouping	.23

The conventional syllogisms in False Premises and Reasoning I clearly stamp them as tests of deduction. In the PMA study (6), these two tests were the highest on the deductive factor. Apparently, Figure Classification, so different in content from the syllogistic tests, involves the same factor. In this test, having grasped the principle of construction in a problem easily, the subject has then to apply the principle in a deductive manner as he designates the items in the trial group that belong to the first of the standard groups. Similarly in the Verbal Analogies tests, the principal reasoning component would

seem to be, not the induction of the principle involved in the ratio, but the deductive element concerned with selecting the element which bears out the principle already grasped. The loading of Figure Grouping on this factor is probably due to the very high correlation between this test and its parent test Figure Classification.

The interpretation of this reasoning factor is straightforward. A point might be raised, however, concerning the discrepancies between the deductive factor in the Hyde Park study (8) and this present factor. In that battery the order of loadings on deduction was: Arithmetic, Number Series, Mechanical Movements, Reasoning II, Reasoning III, Verbal Analogies, and Reasoning I. Many of these tests in the present battery have shifted over to the inductive factor. But in the previous study the subjects were younger and likewise a less selected group. The PMA study (6), which also used college students as subjects, gives a factorial pattern for the reasoning tests that is more consonant with the results of the present study. Perhaps the younger, less selected, subjects are forced to use more deduction in the solution of these tests, whereas for the more able and more experienced groups, no analytic procedures are necessary. They would tend to adopt a more synthetic and almost preceptual application of the principles induced. An interesting series of studies might be made of the changes in factorial composition of the same group of tests when given to different age groups.

The observations concerning low loadings and the factorial complexity of reasoning tests made in connection with the induction factor are also pertinent here. The media for the deductive factor are words and figures.

The small number of tests with high loadings on this factor indicate the desirability of another study with a larger number of tests designed to establish more convincingly this deductive factor. One might experiment with various types of syllogisms, including, of course, implicit syllogisms, and formal syllogisms with implied premises.

Factor C: Flexibility of Closure

20.	Copying	.45
21.	Gottschaldt Figures	.41
22.	Designs	.38
23.	Block Counting	.29
24.	Paper Puzzles	.27
12.	Marks	.23
25.	Mechanical Movements	.22
26.	Hidden Words	.22
31.	Four-Letter Words	.20

Tests with high loadings on this factor seem to involve the holding in mind of a configuration or gestalt, and the operating with it against distractors. Thus in Copying the subject must keep in mind the figure he is trying to reproduce, and not allow the regular pattern of dots to distract him as he connects the appropriate dots. In Gottschaldt Figures the subject must hold in mind the standard figure as he decides whether or not it is embedded in the more complex trial figures. That subject does best who has a clear image of the standard, and need not refer to it often while examining the trial figures. Similarly in Designs, the subject retains in mind the image of the capital Sigma, while he examines the trial figures. In Block Counting, the subject keeps in mind the entire formation of the pile of blocks as well as their sizes and shapes, as he counts those which touch the block in question. In Paper Puzzles the subject must keep in mind the size and shape of the large figure into which the smaller pieces must fit. In Marks, the factor is less evident, but perhaps the subject is aided by keeping in mind the spatial location of the Marks in the preceding parts of an item, as he tries to verify the proposed solution in later lines. In Mechanical Movements, again, the subject may be helped if he can recall and visualize the spatial relations and the form of the pulleys, gears, etc., as he answers the verbal questions. In Hidden Words, if one keeps the pattern of four-letters-in-a-straight-line in mind, one finds it easier to operate amid the confusion of letters. Similarly in Four-Letter Words, it helps in the spotting of the four-letter words, if one has formed some sort of image of a group of four-letter words, and holds this against the distractions of the line of letters.

This interpretation of factor C is strengthened by Thurstone's interpretation of a similar factor in his recent study of mechanical aptitude (12). The following tests have loadings on this factor called by him the Second Closure Factor, C_2, "flexibility of closure:"

Designs	.38
Copying	.36
Paper Puzzles	.32
Gottschaldt Figures	.30
Block Counting	.20
Mechanical Movements	.11

These tests correspond quite nicely to the list of tests in the present battery on factor C. The few tests in this battery not found on the list above were not included in Thurstone's battery. Similar factors have been reported by Rimoldi (4) and Yela (14). Here then we

would seem to have a stable ability, whose importance in mental life has yet to be determined.

It is interesting to note the relationship between Thurstone's restricted thinking factor in the PMA study (6) and this flexibility of closure factor. Copying, Block Counting, Form Board (similar to Paper Puzzles), and Mechanical Movements were four of the six tests with highest loadings on the restricted thinking factor. The other tests with loadings on the closure factor were not included in the former study. Perhaps these two factors are really the same.

Factor D: Speed of Closure

27.	Street Gestalt	.49
28.	Backward Writing	.46
29.	Mutilated Words	.41
30.	Incomplete Words	.37
35.	Hidden Pictures	.34
31.	Four-Letter Words	.34
32.	Scrambled Words	.26
36.	Identical Forms	.26
34.	Picture Squares	.23
33.	Hidden Letters	.22

This factor should probably be identified with what Thurstone has called recently (12), the first Closure Factor, C_1, "speed of closure," and what he termed speed of perception in the *Factorial Study of Perception* (11). It is likewise similar to Bechtoldt's factor G, which he described as "facility in restructuring formal perceptual material possessing a weak intrinsic structure" (1). Meili (3) describes a factor similar to the present factor, which he calls "globalization," a facility for combining distinct elements to form a whole. Street Gestalt had a prominent loading on "globalization." The factor clearly involves something more than mere speed of perception, such as you might find in a cancellation of letters test. In all of the tests with loadings on this factor, there is an unstructured field, in which some reorganization must occur. This process of reorganization may well be termed closure.

In Street Gestalt, the subject must reconstruct the original picture. In Backward Writing, he must reverse the already reversed words. In Mutilated Words he must reconstruct the original words whose parts have been erased. In Incomplete Words, the missing letters must be supplied. In Hidden Pictures the often poorly structured elements of the hidden faces, etc., must be fused into wholes. In Four-Letter Words, the spaced letters of the words must be fused. In Scrambled Words the whole word must be reassembled. In Hidden

Letters, the dots must be united to form the letter or digit demanded. It is more difficult to verify this interpretation in Identical Forms and Picture Squares. However, in both tests there is the perception of rather complex detail and the identification of this with something already seen. The process of identification may be similar to the structuring of formal visual material.

In both of Thurstone's studies referred to above, Street Gestalt and Mutilated Words have high loadings on this factor of speed of closure. In Bechtoldt's study, Four-Letter Words, Mutilated Words, and Hidden Pictures, were prominent on factor G. Their presence also on the speed of closure factor of the present study leads to the tentative identification of factor G with speed of closure.

Factor E: Space

37.	Figures	.46
38.	Cards	.43
39.	Solid Blocks	.40
19.	Figure Grouping	.35
28.	Backward Writing	.32
6.	Pattern Analogies	.31
36.	Identical Forms	.27
15.	Figure Classification	.26
17.	Verbal Analogies I	.22
34.	Picture Squares	.22
25.	Mechanical Movements	.21

The three highest tests on this factor unmistakably stamp it as a space factor, since they are standard tests of space. In his recent study on mechanical aptitude (12), Thurstone isolated three space factors. The first of them has the same three tests in identical order at the head of the factor loadings.

Backward Writing can be solved by revolving mentally the reversed word in space. Identical Forms, Figure Classification, and Mechanical Movements had similar loadings on the space factor in the PMA study (6). Figure Grouping is a variant of Figure Classification, and like Pattern Analogies, seems to involve the comparison and moving around in space of geometrical forms.

Factor F: Verbal Comprehension

40.	Definition	.71
41.	Vocabulary	.62
42.	Completion	.52
17.	Verbal Analogies I	.39
18.	Verbal Analogies II	.30
5.	Reasoning II	.28
7.	Reasoning III	.28

16.	Reasoning I	.25
11.	Tabular Completion	.20

Again we have a stable Primary Mental Ability, whose nature can be determined from the tests with highest loadings. Definition and Vocabulary are standard tests for the verbal comprehension factor. Completion, a form of sentence completion test, failed to show any significant loadings on the closure factors, but proved to be a good test of verbal comprehension. This is not surprising, since the greater one's comprehension of the words in the sentences with all their connotations, the more rapid the completion of the sentences by the subject. Verbal Analogies tests have always involved a great deal of verbal comprehension, for the obvious reason that one cannot complete the analogy unless the words in the first ratio are clearly understood and their logical verbal relationship comprehended. Reasoning I, II, and III, are completely verbal in content. Tabular Completion also had a small loading on this factor in the PMA study (6). Apparently the verbal headings on the columns and rows of the tables of this test introduce a verbal component; Tabular Squares has no verbal headings, and no loading on the present factor.

There is also an interesting difference in the loadings of Reasoning I and False Premises on this factor. The syllogisms in Reasoning I are conventional in content as well as in form, but False Premises employs utterly ridiculous premises and meaningless conclusions. In the latter test, consequently, no premium is placed upon understanding the premises and conclusions. The subject is forced to concentrate almost exclusively upon the deductive reasoning involved in the test. In the oblique solution, therefore, Reasoning I has a moderate loading on Verbal Comprehension, False Premises a loading of only .032.

Factor G: Word Fluency

43.	First Letter	.59
30.	Incomplete words	.48
44.	Suffixes	.48
32.	Scrambled Words	.43
3.	Letter Grouping	.25
1.	Letter Series	.24
28.	Backward Writing	.23
31.	Four-Letter Words	.22
29.	Mutilated Words	.21

In all of these tests one detects the operation of the ability to think of words rapidly, which characterizes the word fluency factor. First Letter and Suffixes are conventional tests of this ability. Sub-

jects with the ability to think of words rapidly obviously do better on Incomplete Words, in which they are obliged to supply the missing letters in words; on Scrambled Words, in which they must reassemble four-letter words; on Backward Writing, in which they must reverse words; on Four-Letter Words, in which they must spot words in a line of evenly spaced letters; on Mutilated Words, in which they must fill in the erasures of the maimed words. It is somewhat surprising to find Letter Grouping and Letters Series with even such small loadings on word fluency, but this is not without precedent, since they had even higher loadings on this factor in the Hyde Park study (8).

Factor H: Number

45.	Multiplication	.65
46.	Addition	.54
32.	Scrambled Words	.33
28.	Backward Writing	.28
13.	Numerical Judgment	.27
31.	Four-Letter Words	.27
10.	Tabular Squares	.24
30.	Incomplete Words	.24
4.	Arithmetic	.23

The ability to perform simple numerical operations clearly defines this factor. Multiplication and Addition are stock tests for this factor. Nor is it unusual to find reasoning tests such as Numerical Judgment, Tabular Squares, and Arithmetic with a number component.

It is, however, interesting to note the presence of Scrambled Words, Backward Writing, Four-Letter Words, and Incomplete Words, with even moderate loadings on the number factor. This is in line with the hypothesis of Landahl and Coombs (quoted by Thurstone, 10, pp. 199-200) that the number factor really measures "facility with highly practiced associations." Certainly all of these verbal, non-numerical tests operate with very well known, and hence well-practiced, words. The loadings are not large enough to substantiate the hypothesis but they are suggestive.

Factor J

34.	Picture Squares	.36
35.	Hidden Pictures	.34
36.	Identical Forms	.26
7.	Reasoning III	.23
12.	Marks	.22

8.	Secret Writing	.22
26.	Hidden Words	.21

It is difficult to determine whether this factor is residual or allied to Bechtoldt's factor Y (1). Picture Squares and Identical Forms were included in the present battery precisely because they showed loadings on Bechtoldt's factor Y. But all the loadings on this present factor are small, and Hidden Pictures, the second highest test on this factor, had a loading of zero on Bechtoldt's Y. If this factor were to be identified with Bechtoldt's, however, the change in the factorial composition of Hidden Pictures might be due to alterations in time limits and instructions. The time limits were lengthened for this battery. And instead of allowing the subjects to work as long as they wished on an individual picture, they were instructed to find a limited number of relatively easy pictures in each problem first, and then after finishing all of the problems to come back and look for the more difficult hidden pictures. This change in instructions probably increased the scores somewhat. Tentatively, then, the factor might be considered as akin to Bechtoldt's which he described as facility in organizing simultaneous visual configurations under the distraction of continued activity. Thus, in Picture Squares, one must scan the various pictures of the square, trying to pick out two that are identical despite the distraction of the other similar pictures. In Identical Forms, one tries to pick out the form that is identical with the standard, while being distracted by the nearly identical other forms. In Hidden Pictures, one is looking for a person or a face, despite the distraction of the picture as a whole.

It must be admitted, however, that it is difficult to verify this interpretation of the factor in the other tests with loadings over .20. Hence, the factor should be considered as a residual factor, until this tentative interpretation is substantiated by further investigations in this domain.

Factor K: Residual

36.	Identical Forms	.38
3.	Letter Grouping	.34
15.	Figure Classification	.27
19.	Figure Grouping	.26
1.	Letter Series	.25
33.	Hidden Letters	.23

No interpretation has been made of this factor. There is nothing which the above tests seem to have in common. Hence it has been considered a residual factor.

The Second-Order Domain

Table 6 reproduces the correlations between the primary factors. The intercorrelations between the eight clearly interpretable primaries are positive for the most part except for the speed of closure factor, which has a large negative correlation with the number factor, and small negative correlations with several other factors. The high correlation between space and the second closure factor, flexibility of closure, is not surprising when we consider that all of the tests with high loadings on the closure factor have shown fairly high loadings on the space factor in other studies, and that in Thurstone's *Factorial Study of Perception* (10), the composite space test and Gottschaldt Figures, one of the primary tests of flexibility of closure, both had significant loadings on the same factors, A and E.

In factoring Table 6 the last two factors were neglected—the last factor because it seems to be residual, and the other factor because its interpretation is not definite and the loadings on it are low. The complete centroid method for factoring was employed again, and a number of trials were made until the communalities were very stable —varying but a few thousandths on the last two runs—and the residuals uniformly small. The resulting orthogonal factor matrix F_2 is reproduced in Table 7.

The matrix F_2 was then rotated to an oblique simple structure. The resulting oblique factor matrix V_2 and its transformation matrix are reproduced in Tables 8 and 9 respectively. Table 10 gives the cosines between the reference vectors.

Since there are four factors in the second order and only eight primary factors, it is obviously not possible to determine these four factors with great confidence. The interpretation, then, of the second order can only be tentative. Such interpretations, however, are often interesting and meaningful. Final judgment must be reserved until the results of this second order are confirmed by succeeding studies.

Factor α

Space	.74
Deduction	.68
Induction	.67
Flexibility of Closure	.64

The loadings on this factor remind one of the findings in the *Factorial Study of Perception* (10). There Thurstone found that the composite reasoning test had a high loading on his factor E, flexibility of closure. There was also a correlation of .39 between the composite reasoning test and Factor A, "strength of configuration," which included, among the tests with loadings on it, the space composite and

Gottschaldt Figures. Since space also had a loading on factor E of .22, it is not surprising that space should be represented on this present second-order factor. It has been suggested (4) that Thurstone's A and E are closely related and may sometimes be fused into a single factor. At any rate Thurstone's results seem in line with those presented here.

Recently Yela, in refactoring some of Alexander's data (14), reported a correlation of .59 between a reasoning factor and a perceptual factor that he identified with the flexibility of closure factor. He also found space, the closure factor, and the reasoning factor on a second-order factor.

Thurstone, again, in his most recent study on mechanical aptitude, got a correlation of .63 between induction and flexibility of closure, .38 between induction and space, and .53 between space and flexibility of closure. All of these results are in harmony with those of the present study.

One explanation of this factor may be the fact that it is possible to solve the space and closure tests in either one of two ways, analytically or synthetically. In the analytic solution, one analyzes the problem and arrives at a solution by a process of logical reasoning. For example, one traces out the standard figures in the trial figures of Gottschaldt Figures, one compares the length of lines and the size of the angles in Copying, one reasons to the true position of the cut corner when the forms are turned in Cards. In the synthetic procedure, one actually sees the standard figure in the trial figures in Gottschaldt Figures, one traces out a pattern on the dots with the image of the pattern clearly in mind in Copying, and in Cards, Figures, and Solid Blocks, one imagines the rotation of the card, figure, or block. Apparently the synthetic process is more effective in solving the space and closure tasks, and the first-order factors reflect this process. But since the synthetic and analytic procedures are not entirely opposed, the analytic procedure might be reflected in this second-order factor and in the correlations between the primaries.

Another explanation emphasizes the fact that there are certain configurational or gestalt elements in both induction and deduction. In searching for a principle in a particular item, one must keep in mind the elements of the problem, and the relationships between these elements may be visualized in a spatial and configurational manner. In deduction one may solve the syllogisms with the assistance of a spatial framework or configuration, in which the middle term is the bond or link between the other two terms. Or in the analogies test, one may represent the proportions in a spatial arrangement. Or in

the application of a test like Figure Classification, a person might keep some general examples of the rule governing an item in mind while he marks the test symbols.

Whatever be the true explanation for the relationship between the reasoning factors and the flexibility of closure factor, the fact itself seems undeniable. The present study has confirmed its existence more strongly by demonstrating its presence even when the reasoning and closure factors have been adequately determined in the same battery, which has not been the case in the studies hitherto.

Factor β

Number	.58
Word Fluency	.57
Verbal Comprehension	.49
Speed of Closure	—.46

This is a bipolar factor. Analysis of the primaries at the extremes leads to the tentative interpretation of this factor as a sort of speed of association factor. At one extreme are well-practiced, drilled, common associations such as numbers, words beginning with certain letters or involving frequently used words, and even the practiced association of particular connotations or meanings with particular words such as one finds in the tests of verbal comprehension. At the other pole is the speed of closure factor, in which one is required to fill in or complete unstructured configurations. Such a task is unusual and not commonly experienced, even though the figures and words themselves are not uncommon. Individuals who are adept in working with the mechanical sort of tasks at the positive pole, would, under this interpretation, find difficulty in the more imaginative unfamiliar tasks required in the closure tests.

It is interesting to note that Street Gestalt, which has the highest loading on the speed of closure factor, had a negative loading of .25 on the number factor in Bechtoldt's study (1); that the correlation between the composite number test and the composite speed of closure test in the *Factorial Study of Perception* (10) was —.15; that the perceptual primary in Taylor's study (5) had a negative correlation of .24 with number. It would be interesting to investigate this phenomenon in a further study.

Factor γ

Deduction	.77
Verbal Comprehension	.57
Induction	.39

Tentatively this second-order factor may be interpreted as fa-

cility in handling meaningful verbal materials. An alternative interpretation might look upon it as an analytic ability, something akin to Spearman's noegenesis, the ability to grasp and discover relations. Finally it could be considered as the ability to do abstract thinking. All of the primaries with loadings on this second-order factor are concerned with meaningful situations, and also with verbal material to some degree at least. In deductive tests like the syllogism tests and the verbal analogies, one is concerned with meaningful verbal materials. The verbal analogies tests and the reasoning tests had loadings on verbal comprehension also. In addition, Definition involves the understanding of the meaning of definition of a word. Vocabulary requires the selection of a synonym for a word in a meaningful phrase. Completion demands an understanding of a sentence and the reality corresponding to the sentence, before the subject can supply the correct words to complete the sentence. Finally, in all tests of induction one is concerned with a rule or principle underlying a meaningful arrangement of verbal, numerical, or pictorial materials, and the rule is often phrased by the subject verbally to himself in the course of solving an item. Furthermore, you have meaningful verbal materials in Arithmetic, Reasoning II, Reasoning III, and Verbal Analogies III.

It is significant, too, that the space and closure primaries, which are primarily synthetic in character are not found on this factor. Nor does the number factor, which is concerned with highly practiced, mechanical associations, have a loading on this factor. Word fluency is absent too, but this factor calls for a spontaneous and somewhat mechanical recall of words. One recalls the words not because of their connotations but rather because of the positions of individual letters or groups of letters in the words.

Factor 8

Flexibility of Closure	.56
Speed of Closure	.40
Space	.33
Word Fluency	.26

One might call this factor a configurational or perhaps a closure factor. Thurstone's factor A, strength of a configuration, had among the tests with loadings on it, the composite space test, Gottschaldt Figures which has a high loading on the flexibility of closure factor, and Street Gestalt and Mutilated Words, which are prominent on the speed of closure factor. In addition the composite word fluency test had a correlation of .20 with this factor. Perhaps this present factor is to be identified with factor A, which has not shown up elsewhere in this battery.

TABLE 1
Product-Moment Correlation Coefficients
(With Decimal Points Omitted)

	Rel.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1.	86		65	63	55	47	53	57	47	54	47	44	43	44	38	45
2.	77	65		54	51	53	55	55	43	60	43	45	36	43	33	46
3.	92	63	54		57	45	53	55	40	51	41	39	41	36	36	41
4.	87	55	51	57		39	51	50	43	50	48	46	38	38	34	40
5.	62	47	53	45	39		49	56	40	59	39	45	33	28	36	44
6.	75	53	55	53	51	49		55	43	54	38	41	34	35	33	50
7.	79	57	55	55	50	56	55		53	60	40	47	40	37	45	42
8.	97	47	43	40	43	40	43	53		53	35	38	41	40	25	44
9.	78	54	60	51	50	59	54	60	53		48	45	32	52	42	48
10.	96	47	43	41	48	39	38	40	35	48		38	31	35	25	37
11.	92	44	45	39	46	45	41	47	38	45	38		33	33	27	24
12.	82	43	36	41	38	33	34	40	41	32	31	33		26	21	36
13.	65	44	43	36	38	28	35	37	40	52	35	33	26		22	34
14.	45	38	33	36	34	36	33	45	25	42	25	27	21	22		44
15.	98	45	46	41	40	44	50	42	44	48	37	24	36	34	44	
16.	69	35	43	28	32	51	34	58	36	50	23	36	23	23	51	49
17.	92	49	42	46	39	49	45	55	37	54	26	34	25	31	42	46
18.	86	49	51	48	40	56	47	57	45	57	28	36	38	37	43	52
19.	88	48	44	49	46	40	54	46	45	49	34	29	35	37	38	70
20.	92	40	39	44	46	28	41	41	37	34	32	28	46	29	30	47
21.	81	38	37	41	41	33	49	44	43	39	29	32	37	41	35	53
22.	94	27	24	29	36	23	39	33	36	27	25	35	26	18	24	31
23.	97	48	43	51	50	34	45	45	41	45	39	36	46	38	35	49
24.	82	48	47	42	45	39	50	36	41	42	32	34	40	35	26	45
25.	88	38	43	36	47	44	54	45	40	51	20	39	37	44	44	54
26.	89	41	44	43	46	39	46	45	45	41	35	36	32	34	25	46
27.	70	09	03	06	03	–01	14	09	08	–03	–06	–05	09	07	07	18
28.	97	45	43	48	36	31	51	38	36	34	35	35	25	30	19	40
29.	82	29	17	18	18	17	21	22	21	20	10	17	17	09	20	19
30.	91	37	38	41	33	27	31	34	36	33	35	25	20	19	15	23
31.	88	27	23	33	31	11	14	24	29	21	26	17	20	13	14	21
32.	93	40	36	47	46	29	30	38	32	39	45	29	25	24	23	24
33.	79	23	11	23	29	11	29	16	19	18	12	19	20	16	16	31
34.	83	25	23	29	31	21	39	33	31	26	16	16	27	18	12	20
35.	69	06	03	06	10	–08	13	13	08	–08	–04	01	09	00	04	07
36.	98	41	22	47	35	18	40	35	30	22	25	27	34	25	20	37
37.	96	39	32	34	36	23	46	33	36	35	26	29	36	23	23	44
38.	96	47	40	42	44	33	51	36	39	36	38	36	39	34	31	50
39.	74	33	36	33	36	38	48	35	33	40	25	36	31	31	24	43
40.	90	24	31	32	22	49	31	48	27	44	19	29	05	15	30	30
41.	96	24	33	31	20	42	23	48	26	42	25	29	13	07	26	28
42.	81	44	38	45	40	52	36	61	44	56	31	47	25	34	39	41
43.		16	18	32	13	20	18	19	12	15	20	06	10	01	09	12
44.		15	12	20	07	14	10	08	03	17	14	11	08	01	13	–01
45.	89	33	34	37	40	21	23	33	30	43	44	37	22	39	14	20
46.	90	32	34	31	29	18	17	25	19	35	34	38	21	26	10	12

TABLE 1 (Continued)

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
35	49	49	48	40	38	27	48	48	38	41	09	45	29	37
43	42	51	44	39	37	24	43	47	43	44	03	43	17	38
28	46	48	49	44	41	29	51	42	36	43	06	48	18	41
32	39	40	46	46	41	36	50	45	47	46	03	36	18	33
51	49	56	40	28	33	23	34	39	44	39	–01	31	17	27
34	45	47	54	41	49	39	45	50	54	46	14	51	21	31
58	55	57	46	41	44	33	45	36	45	45	09	38	22	34
36	37	45	45	37	43	36	41	41	40	45	08	36	21	36
50	54	57	49	34	39	27	45	42	51	41	–03	34	20	33
23	26	28	34	32	29	25	39	32	20	35	–06	35	10	35
36	34	36	29	28	32	35	36	34	39	36	–05	35	17	25
23	25	38	35	46	37	26	46	40	37	32	09	25	17	20
23	31	37	37	29	41	18	38	35	44	34	07	30	09	19
51	42	43	38	30	35	24	35	26	44	25	07	19	20	15
49	46	52	70	47	53	31	49	45	54	46	18	40	19	23
	48	56	42	31	32	22	26	26	44	33	–02	20	22	25
48		67	49	34	35	22	34	38	41	25	13	33	23	23
56	67		50	40	38	25	39	40	49	45	08	29	17	28
42	49	50		47	46	38	53	52	54	48	19	45	24	23
31	34	40	47		65	48	54	60	52	51	20	37	24	34
32	35	38	46	65		56	61	59	58	46	25	45	22	30
22	22	25	38	48	56		51	41	40	42	25	42	24	26
26	34	39	53	54	61	51		61	54	55	21	39	18	29
26	38	40	52	60	59	41	61		58	41	20	37	21	21
44	41	49	54	52	58	40	54	58		42	21	34	18	23
33	25	45	48	51	46	42	55	41	42		13	38	24	37
–02	13	08	19	20	25	25	21	20	21	13		27	33	14
20	33	29	45	37	45	42	39	37	34	38	27		30	48
22	23	17	24	24	22	24	18	21	18	24	33	30		41
25	23	28	23	34	30	26	29	21	23	37	14	48	41	
06	16	19	23	28	32	32	31	15	12	27	19	45	22	55
20	34	36	26	33	34	35	35	19	27	40	10	50	25	64
04	14	17	31	25	22	20	29	28	35	24	24	29	11	13
15	22	26	36	30	23	26	41	30	31	35	20	30	26	21
–02	05	00	05	19	22	19	15	09	09	24	30	21	22	15
08	35	25	43	51	44	41	48	39	29	34	27	51	25	26
18	30	29	53	38	39	44	57	48	43	34	11	36	13	14
18	35	37	53	51	54	41	65	61	54	46	17	49	22	28
21	28	32	48	42	49	39	59	48	54	42	09	35	11	21
48	58	55	29	25	25	22	18	23	31	18	01	25	14	25
46	51	50	25	23	12	19	12	12	22	18	–05	19	10	31
53	64	56	44	33	31	25	32	29	34	30	06	35	20	33
11	24	24	17	23	13	17	12	11	10	17	06	28	11	51
10	20	20	05	14	05	01	05	11	01	11	01	21	18	29
12	24	21	23	30	26	23	30	18	13	27	–11	36	–03	29
08	20	17	10	15	13	12	20	07	02	17	–11	32	02	35

TABLE 1 (Continued)

31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
27	40	23	25	06	41	39	47	33	24	24	44	16	15	33	32
23	38	11	23	03	22	32	40	36	31	33	38	18	12	34	34
33	47	23	29	06	47	34	42	33	32	31	45	32	20	37	31
31	46	29	31	10	35	36	44	36	22	20	40	13	07	40	29
11	29	11	21	–08	18	23	33	38	49	42	52	20	14	21	18
14	30	29	39	13	40	46	51	48	31	23	36	18	10	23	17
24	38	16	33	13	35	33	36	35	48	48	61	19	08	33	25
29	32	19	31	08	30	36	39	33	27	26	44	12	03	30	19
21	39	18	26	–08	22	35	36	40	44	42	56	15	17	43	35
26	45	12	16	–04	25	26	38	25	19	25	31	20	14	44	34
17	29	19	16	01	27	29	36	36	29	29	47	06	11	37	38
20	25	20	27	09	34	36	39	31	05	13	25	10	08	22	21
13	24	16	18	00	25	23	34	31	15	07	34	01	01	39	26
14	23	16	12	04	20	23	31	24	30	26	39	09	13	14	10
21	24	31	20	07	37	44	50	43	30	28	41	12	–01	20	12
06	20	04	15	–02	08	18	18	21	48	46	53	11	10	12	08
16	34	14	22	05	35	30	35	28	58	51	64	24	20	24	20
19	36	17	26	00	25	29	37	32	55	50	56	24	20	21	17
23	26	31	36	05	43	53	53	48	29	25	44	17	05	23	10
28	33	25	30	19	51	38	51	42	25	23	33	23	14	30	15
32	34	22	23	22	44	39	54	49	25	12	31	13	05	26	13
32	35	20	26	19	41	44	41	39	22	19	25	17	01	23	12
31	35	29	41	15	48	57	65	59	18	12	32	12	05	30	20
15	19	28	30	09	39	48	61	48	23	12	29	11	11	18	07
12	27	35	31	09	29	43	54	54	31	22	34	10	01	13	02
27	40	24	35	24	34	34	46	42	18	18	30	17	11	27	17
19	10	24	20	30	27	11	17	09	01	–05	06	06	01	–11	–11
45	50	29	30	21	51	36	49	35	25	19	35	28	21	36	32
22	25	11	26	22	25	13	22	11	14	10	20	11	18	–03	02
55	64	13	21	15	26	14	28	21	25	31	33	51	29	29	35
	52	15	20	19	36	15	21	15	21	32	30	37	21	32	31
52		06	26	10	24	23	35	24	27	34	36	50	26	45	44
15	06		18	10	33	25	26	31	04	–03	20	00	00	00	–06
20	26	18		24	43	32	39	37	14	08	24	12	05	20	11
19	10	10	24		24	15	23	03	–07	–07	01	07	05	–07	–02
36	24	33	43	24		46	54	37	19	11	30	14	14	25	14
15	23	25	32	15	46		70	58	11	06	18	07	09	25	18
21	35	26	39	23	54	70		61	13	07	25	10	12	32	19
15	24	31	37	03	39	58	61		10	01	24	09	07	27	13
21	27	04	14	–07	19	11	13	10		76	62	34	18	18	17
32	34	–03	08	–07	11	06	07	01	76		60	39	15	23	30
30	36	20	24	01	30	18	25	24	62	60		22	22	34	28
37	50	00	12	07	14	07	10	09	34	39	22		48	20	28
21	26	00	05	05	14	09	12	07	18	15	22	48		14	22
32	45	00	20	–07	25	25	32	27	18	23	34	20	14		69
31	44	–06	11	–02	14	18	19	13	17	30	28	28	22	69	

TABLE 2
Centroid Factor Matrix *F*

	I	II	III	IV	V	VI	VII	VIII	IX	X	h^2
1.	.710	–.111	.058	–.198	.085	.255	.107	–.040	–.193	–.011	.6815
2.	.676	–.184	.135	–.225	.110	.086	.123	–.075	–.023	–.139	.6197
3.	.702	–.120	–.030	–.104	–.036	.156	.031	.099	–.259	–.056	.6255
4.	.672	–.028	.010	–.225	.132	.038	–.018	.063	–.087	–.051	.5364
5.	.613	–.237	.292	–.027	–.028	–.011	.150	.056	.061	–.132	.5656
6.	.699	.112	.138	–.087	–.088	.118	.118	–.028	.051	–.187	.6016
7.	.730	–.200	.183	.024	.094	.098	.035	.106	.137	.058	.6600
8.	.623	–.014	.105	–.083	.139	.152	–.063	.075	.121	.057	.4760
9.	.714	–.282	.257	–.149	.023	.012	–.028	–.105	.062	.015	.6939
10.	.540	–.185	–.062	–.303	.056	–.018	–.113	–.004	–.133	–.041	.4570
11.	.562	–.148	.103	–.193	.037	–.079	–.063	.077	.156	–.076	.4331
12.	.518	.104	.040	–.160	.084	.019	.023	.186	–.124	.134	.3823
13.	.499	–.034	.138	–.228	.109	.107	–.169	–.165	.009	.072	.4056
14.	.495	–.087	.228	.182	.107	–.093	.087	–.162	–.135	.138	.4287
15.	.665	.160	.272	.126	–.050	.048	–.105	–.266	–.203	.070	.6904
16.	.528	–.245	.381	.269	.155	–.062	.103	–.072	.067	.149	.6267
17.	.645	–.225	.235	.248	–.206	.085	.045	–.034	.022	.090	.6447
18.	.679	–.203	.282	.150	–.056	.047	.149	.061	.017	.102	.6461
19.	.700	.209	.198	.052	–.169	.125	–.045	–.119	–.095	.039	.6465
20.	.664	.268	–.056	.073	.123	–.177	–.033	.147	–.196	.114	.6416
21.	.670	.303	.021	.067	.218	–.203	–.135	–.024	–.025	.015	.6539
22.	.547	.273	–.098	.052	.123	–.241	–.174	.075	.126	–.092	.5205
23.	.708	.312	.020	–.179	.057	–.125	–.005	.074	–.059	.034	.6599
24.	.641	.305	.158	–.085	.028	–.122	.011	.048	–.145	–.048	.5775
25.	.663	.245	.354	.056	.114	–.164	.037	–.072	.047	–.130	.6936
26.	.635	.135	–.041	–.074	.140	.011	.015	.119	.059	.037	.4675
27.	.206	.387	–.165	.334	.099	.192	–.007	–.075	.066	–.033	.3887
28.	.644	.102	–.310	.034	–.108	.126	–.077	–.188	.044	–.210	.6372
29.	.352	.096	–.189	.228	.215	.168	.164	–.088	.127	.006	.3459
30.	.547	–.190	–.475	.088	.173	.034	.130	–.090	.037	–.162	.6523
31.	.448	–.076	–.459	.139	.077	.032	–.201	.041	–.063	–.072	.4946
32.	.594	–.242	–.449	–.029	.078	–.059	.060	–.098	.044	–.117	.6522
33.	.321	.291	.065	.042	–.018	.196	–.120	.045	–.064	–.175	.2835
34.	.445	.229	–.103	–.051	–.089	.162	.102	.138	.234	.082	.3886
35.	.163	.306	–.267	.143	.124	.148	.114	.064	.153	.139	.3090
36.	.573	.333	–.219	.057	–.219	.158	–.139	.150	–.087	.096	.6220
37.	.562	.374	.064	–.201	–.259	–.124	.068	–.050	.027	.062	.5943
38.	.682	.393	–.018	–.211	–.159	–.146	.120	–.098	.018	.067	.7397
39.	.574	.331	.136	–.220	–.169	–.171	.071	–.026	.074	–.019	.5754
40.	.493	–.402	.177	.428	–.225	–.149	–.091	.147	.144	–.105	.7534
41.	.455	–.524	.061	.366	–.135	–.135	–.102	.171	.085	–.058	.7058
42.	.649	–.365	.156	.244	–.092	.061	–.181	.133	.084	.090	.7161
43.	.348	–.268	–.436	.232	–.186	–.190	.245	.056	–.151	–.140	.6130
44.	.233	–.209	–.303	.112	–.162	–.100	.298	.006	–.115	.064	.3446
45.	.475	–.292	–.265	–.406	–.135	–.150	–.353	–.126	.069	.210	.7760
46.	.380	–.394	–.318	–.324	–.115	–.108	–.183	–.122	.059	.093	.5911

TABLE 3
Transformation Matrix

	A	B	C	D	E	F	G	H	J	K
I	.2313	.1525	.1609	.1739	.1847	.1737	.1608	.1502	.1198	.1231
II	–.1865	–.0373	.1385	.2295	.3582	–.3138	–.2177	–.2588	.0891	.0563
III	.2399	.2008	–.0328	–.4105	.0594	.2299	–.4822	–.3352	–.1236	–.0355
IV	–.4357	.2938	.0909	.3786	.0201	.4263	.0926	–.2317	.0286	.0937
V	.2670	.2267	.5205	.1232	–.7795	–.4282	.0051	.0550	.0474	–.2503
VI	.5353	.0324	–.5387	.4910	–.0389	–.0710	–.0546	–.0598	.3576	.4622
VII	.1733	.0711	–.2448	–.1910	.1135	–.2620	.6577	–.5416	.1786	–.3650
VIII	.3456	–.5395	.5063	–.3526	–.3567	.4077	–.2152	–.5740	.5149	.1780
IX	–.0632	–.2783	–.2610	.3933	.2232	.4185	–.4209	.2850	.3102	–.7176
X	–.3911	.6501	.0015	–.1830	–.1859	–.1955	–.1634	.1837	.6620	.1246

TABLE 4
Reference Vector Cosines *C*

	A	B	C	D	E	F	G	H	J	K
A	.9999									
B	–.3705	.9999								
C	–.0370	–.0769	.9999							
D	–.0785	.0388	–.3274	1.0000						
E	–.2925	–.1278	–.5725	.2083	1.0000					
F	–.0387	–.3719	–.0079	.0443	.2330	.9998				
G	.0240	.1362	–.1083	–.0006	.0109	–.3777	1.0001			
H	–.2959	.2190	–.1930	.3783	.0427	–.0966	–.1584	1.0000		
J	.1034	.0998	.0073	.0695	–.2208	.0945	–.1892	–.1713	1.0000	
K	.1441	.1540	.0250	–.0176	–.0684	–.0466	.0101	–.1334	.0666	1.0000

TABLE 5
Oblique Factor Matrix V

	I	D	C_2	C_1	S	V	W	N	Y	Res.
1.	.466	.169	–.010	.055	.000	–.090	.241	.059	.089	.248
2.	.448	.064	.007	.010	.035	.012	.188	.101	–.038	.010
3.	.375	.056	.095	.006	.026	.047	.252	.002	.067	.337
4.	.361	.026	.191	.012	–.029	–.023	.113	.101	.059	.118
5.	.348	.024	.021	–.087	.103	.284	.090	–.059	–.003	–.054
6.	.332	–.027	–.050	.146	.309	.126	.099	–.030	.011	.048
7.	.329	.132	.092	.074	–.017	.275	.003	.064	.229	.011
8.	.312	.087	.100	.145	–.017	.118	–.080	.116	.222	.056
9.	.306	.180	–.003	–.030	.070	.185	.014	.225	.000	–.002
10.	.285	.004	.135	–.062	–.053	–.034	.119	.242	–.047	.138
11.	.270	–.078	.124	–.026	.042	.203	–.052	.160	.036	–.091
12.	.236	.079	.236	–.087	–.050	–.047	.031	–.052	.223	.171
13.	.226	.188	.005	.083	.023	–.056	–.096	.292	.012	.090
14.	–.001	.420	.115	–.029	.010	.032	.137	.022	–.018	.049
15.	.022	.415	.024	.127	.261	.020	.012	.097	–.094	.273
16.	.081	.401	.098	–.026	–.056	.251	.014	–.003	.088	–.084
17.	.089	.257	–.099	.070	.220	.391	.070	.020	.113	.145
18.	.213	.340	.021	–.033	.094	.304	.089	–.073	.178	.069
19.	.111	.233	–.039	.155	.353	.103	–.002	.029	.043	.264
20.	.009	.171	.446	.012	.007	–.026	.103	–.052	.139	.186
21.	–.012	.178	.405	.152	.058	–.012	–.026	.106	.010	.021
22.	–.042	–.058	.377	.188	.103	.119	–.074	.124	.009	–.077
23.	.152	.055	.291	.010	.154	–.062	.027	.021	.104	.077
24.	.155	.080	.272	–.039	.179	–.022	.025	–.091	–.010	.122
25.	.140	.144	.224	.042	.214	.122	–.061	–.069	–.096	–.092
26.	.213	.038	.224	.124	.018	.028	.039	.046	.208	.028
27.	–.099	.111	.017	.485	.136	–.032	.039	–.038	.121	.086
28.	.080	–.031	–.081	.461	.324	.063	.232	.276	–.070	.121
29.	.054	.161	.000	.405	–.004	–.046	.212	.039	.178	–.065
30.	.126	.003	.070	.369	–.044	–.017	.481	.240	.011	–.056
31.	–.004	–.027	.198	.339	–.068	.077	.219	.265	.041	.193
32.	.096	–.014	.071	.264	.011	.020	.427	.332	–.017	–.060
33.	.185	–.063	.047	.220	.163	.048	–.083	–.082	–.007	.229
34.	.139	–.070	–.048	.230	.219	.098	–.017	–.013	.355	–.007
35.	–.056	.056	.040	.341	.023	–.099	.068	–.030	.335	–.025
36.	.015	–.001	.084	.260	.271	.094	.018	.021	.262	.382
37.	–.004	.027	.005	–.008	.464	–.010	–.001	.017	.066	.031
38.	.011	.083	.058	.051	.431	–.097	.102	.064	.074	–.007
39.	.069	–.020	.068	–.040	.404	.023	–.033	.001	.016	–.059
40.	–.028	.002	.079	.023	.136	.706	–.007	.003	–.016	.002
41.	.009	.013	.124	–.007	–.025	.624	.068	.070	.014	.029
42.	.131	.147	.080	.080	.007	.521	–.070	.137	.172	.169
43.	–.100	–.066	.067	.049	.100	.159	.591	–.018	–.065	.035
44.	–.089	.067	–.042	–.026	.077	.024	.477	–.024	.070	.015
45.	–.030	.036	.020	–.019	.052	.036	–.031	.653	.024	.061
46.	.024	–.019	–.046	.000	.016	.040	.141	.544	–.019	–.011

TABLE 6

Correlations between Primary Factors, R_{pq}

	I	D	C_2	C_1	S	V	W	N	K	Resid.
I	1.000	.498	.427	–.025	.503	.138	.058	.283	.033	–.197
D	.498	1.000	.262	.051	.278	.321	–.045	–.068	–.138	–.250
C_2	.427	.262	1.000	.133	.672	–.013	.149	.257	.186	–.068
C_1	–.025	.051	.133	1.000	–.091	–.059	–.111	–.382	–.189	–.050
S	.503	.278	.672	–.091	1.000	–.133	.066	.258	.289	–.058
V	.138	.321	–.013	–.059	–.133	1.000	.344	.118	–.101	–.026
W	.058	–.045	.149	–.111	.066	.344	1.000	.311	.241	.038
N	.283	–.068	.257	–.382	.258	.118	.311	1.000	.310	.114
K	.033	–.138	.186	–.189	.289	–.101	.241	.310	1.000	.010
Resid.	–.197	–.250	–.068	–.050	–.058	–.026	.038	.114	.010	1.000

TABLE 7

Centroid Factor Matrix F_2

	I	II	III	IV	h_2
I	.630	.287	.047	–.166	.5090
D	.502	.388	.530	–.255	.7485
C_2	.597	.413	–.207	.345	.6889
C_1	–.227	.396	.275	.337	.3975
S	.642	.443	–.369	.078	.7506
V	.359	–.352	.548	.062	.5569
W	.377	–.436	.056	.363	.4671
N	.531	–.346	–.328	–.089	.5172

TABLE 8

Oblique Factor Matrix V_2

	α	β	γ	δ
I	.678	.112	.393	.053
D	.669	–.043	.771	–.020
C_2	.640	–.011	.001	.563
C_1	.033	–.455	.033	.401
S	.743	–.011	–.019	.333
V	.015	.492	.571	.008
W	–.085	.572	.065	.263
N	.173	.583	–.023	–.099

TABLE 9

Transformation Matrix Λ_2

	α	β	γ	δ
I	.7245	.5566	.4460	.1806
II	.6621	–.8308	.0577	.3259
III			.8248	
IV	–.1917		–.3425	.9280

TABLE 10

Reference Vector Cosines C_2

	α	β	γ	δ
α	1.0000			
β	–.1468	1.0000		
γ	.4270	.2003	.9998	
δ	.1687	–.1702	–.2185	1.0000

REFERENCES

1. Bechtoldt, H. P. Factorial study of perceptual speed. Unpublished Ph.D. dissertation: Department of Psychology, University of Chicago, 1947.
2. Guilford, J. P. (Ed.) Printed classification tests. Report No. 5, AAF Aviation Psychology Program Research Reports. Washington: U. S. Government Printing Office, 1947.
3. Meili, R. L'analyse of l'intelligence. *Arch. Psychol.*, Geneve, 1946, 31, 1-69.
4. Rimoldi, H. J. A. Study of some factors related to intelligence. *Psychometrika*, 1948, 13, 27-47.
5. Taylor, C. S. A factorial study of fluency in writing. Unpublished Ph.D. dissertation: Department of Psychology, University of Chicago, 1946.
6. Thurstone, L. L. Primary mental abilities. Psychometric Monograph No. 1, 1938.
7. Thurstone, L. L. The perceptual factor. *Psychometrika*, 1938, 3, 1-17.
8. Thurstone, L. L. An experimental study of simple structure. *Psychometrika*, 1940, 5, 153-168.
9. Thurstone, L. L. Current issues of factor analysis. *Psychol. Bull.*, 1940, 37, 189-236.
10. Thurstone, L. L. A factorial study of perception. Chicago: Univ. Chicago Press, 1944.
11. Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.
12. Thurstone, L. L. Mechanical aptitude III: analysis of group tests. Chicago: Psychometric Laboratory, Univ. of Chicago, No. 35, 1949.
13. Thurstone, L. L., and Thurstone, T. G. Factorial studies of intelligence. Psychometric Monograph No. 2, 1941.
14. Yela, Mariano. Application of the concept of simple structure to Alexander's data. *Psychometrika*, 1949, 14, 121-135.

Manuscript received 9/13/50
Revised manuscript received 2/5/51

PSYCHOMETRIKA—VOL. 16, NO. 4
DECEMBER, 1951

MATHEMATICAL STRUCTURES AND PSYCHOLOGICAL MEASUREMENTS*

ANDRÉ M. WEITZENHOFFER
UNIVERSITY OF DETROIT

The nature of psychological measurements in relation to mathematical structures and representations is examined. Some very general notions concerning algebras and systems are introduced and applied to physical and number systems, and to measurement theory. It is shown that the classical intensive and extensive dimensions of measurements with their respective ordinal and additive scales are not adequate to describe physical events without the introduction of the notions of dimensional units and of dimensional homogeneity. It is also shown that in the absence of these notions, the resulting systems of magnitudes have only a very restricted kind of isomorphism with the real number system, and hence have little or no mathematical representations. An alternative in the form of an extended theory of measurements is developed. A third dimension of measurement, the supra-extensive dimension, is introduced; and a new scale, the multiplicative scale, is associated with it. It is shown that supra-extensive magnitudes do constitute systems isomorphic with the system of real numbers and that they alone can be given mathematical representations. Physical quantities are supra-extensive magnitudes. In contrast, to date, psychological quantities are either intensive or extensive, but never of the third kind. This, it is felt, is the reason why mathematical representations have been few and without success in psychology as contrasted to the physical sciences. In particular, the Weber-Fechner relation is examined and shown to be invalid in two respects. It is concluded that the construction of multiplicative scales in psychology, or the equivalent use of dimensional analysis, alone will enable the development of fruitful mathematical theories in this area of investigation.

For some years now, I have asked myself a basic question. Why has mathematics been applied so successfully to the physical sciences, but not to psychology? For some peculiar reason, the data of psychology do not appear to be readily amenable to mathematical representation. To date no major mathematical structure has been devised in the domain of psychology. The few attempts in this line have proven themselves rather barren.† Is it in the nature of psychological data

*The editors of this journal should perhaps point out that unanimous agreement with the arguments and points of view expressed in this article is not anticipated. They believe, however, that its publication may stimulate needed thinking and clarification of problems basic to psychological measurement and thus serve the purpose for which the journal was founded.

†Mathematical statistics has of course been very fruitful in dealing with psychological data. This, however, is a matter which is quite different from the main topic of this paper, namely, the mathematical representation of psychological structures.

that one must seek for the answer to this? Or is it in the way these data have been approached and handled?

It is the purpose of this paper to present what is, I believe, a partial answer, if not the whole answer to these questions. It may be stated from the outset that much of the material which follows will appear strange and perhaps forbidding to many readers. The very nature of the problem makes it so. I do not know of any way around this. Other readers may object to the lack of rigor in the treatment which follows, as well as to the brevity in the exposition of many ideas. This too has been largely forced by the necessity of keeping the scope of this paper within limits. An earlier attempt on my part to present a more complete discussion of the same material has shown me that anything really satisfactory would require the writing of a small monograph. Having neither the time nor the inclination to do this at this time, the present material is offered in its stead.

1. Some Basic Notions.

We shall begin by defining an *aggregate*, A, as any collection of specified or unspecified objects one may wish to consider. These objects need not be related to each other in any other way than that of being included in the aggregate. They will be called the elements a, of A.*

It is, however, more usual to deal with aggregates the elements of which have a number of properties in common. We shall call an aggregate a *system*† when all of its elements possess at least one common property besides that of being part of the collection. A system, S, will be considered as given or defined when such a property has been stated, that is, when a necessary and sufficient condition for an object to be an element of S has been given. This will be referred to as the *property of membership*, ε. Any element of a system is called a *member* of it, and this is denoted by

$$a \varepsilon S .$$

An *n-ary relation*, R, in a system S is a property affecting the ordered collection $(a, b, c, \cdots)$ of n elements of S.

*The term aggregate has been used by mathematicians to denote also such entities as classes and sets. This is rather unfortunate as it tends to lead to confusion. As will be seen shortly, the notion as presented here is much more general than that of classes or sets.

†Whether this is to be identified with "set" depends upon the interpretation one places upon the term "rule". As defined by Cantor, a set or Menge is a collection of objects defined by some rule which determines unambiguously which objects belong to the collection and which do not.

A relation of particular importance is that of *equality*, $x = y$, which denotes that the two elements x and y of S are not distinct. The negation of this relation is called *diversity* and is denoted by $x \neq y$.

Another useful relation is that of *partial ordering*. It is a binary relation, $\geqslant$*, such that the following three laws hold for it and any two members of a system,

a. Reflexive Law: $x \geqslant x$ for all $x \varepsilon S$.
b. Anti-symmetric Law: If $x \geqslant y$ and $y \geqslant x$, then $x = y$.
c. Transitive Law: If $x \geqslant y$ and $y \geqslant z$, then $x \geqslant z$.

On the other hand, a relation, $>$, is said to be a *simple ordering* provided:

a. If $x > y$ and $y > z$, then $x > z$.
b. For each pair of elements $(x,y) \varepsilon S$, one of the three properties $x > y$, $x = y$, $y > x$ holds to the exclusion of the other two.

One defines a subsystem, *Sub* S, of S as any system every element of which is an element of S. A subsystem is said to be a *part* of S, or again to be *contained* (or included) in S. This last may be denoted by

$$S' \leq S \text{ (where } S' \text{ denotes } Sub\ S).$$

An *operation*, o, in a system S is a rule which associates with every specified group of elements of S another element of S. It is said to be *n-ary* whenever the specified group consists of n elements.

Clearly, any operation in S is also a relation and hence will partake of all the properties of relations.

Making use of this concept we can now define an *abstract algebra* as any system in which at least one operation is defined. Since algebras are obviously systems too, all of the properties and notions associated with the latter are applicable to algebras within the limits imposed by the operations.

One can go on with this sort of classification. By defining the elements, relations, and operations of abstract algebras one may arrive at the notions of rings, fields, lattices, and so on, which play important roles in mathematics but which are not of interest to us at present.

*Not to be confused with the notion of "greater than or equal to" of ordinary algebra.

It may however be remarked that whether one speaks of systems, algebras, or other mathematical entities, is entirely a question of convenience and of the properties upon which one wishes to focus attention. In other words, it is largely an arbitrary matter, although there are circumstances in which it is more profitable to speak in terms of one rather than another. For sure, classes are both aggregates and systems. They are also abstract algebras and are important enough as such that it is more usual to retain the expression "class" to designate this last aspect. Again, sets are classes, but of a very special kind, being collections of mathematical points.* Thus the use of a special name.

A very important characteristic of systems is their *structure.* This term, which has found its way in nearly every field of human knowledge, remains to date a most elusive notion. Everybody talks about it, but no one defines it. An entire volume has been devoted to the "structure of algebras," yet nowhere in it is it possible to find a definition of structure. Unsatisfactory as it may be, I wish to offer an attempt at some sort of definition of this notion which I believe expresses the consensus of meaning assigned to it. By a *structure* is meant *a totality of relations* conceived as a whole, and independently of the elements between which the relations hold. Thus, the structure of a system such as algebra is the totality of relations holding between its elements, as contrasted to its *content,* which is the totality of its elements, considered independently of any relations between them.

Systems themselves may be related in various ways. Among relations which may hold between systems in general, and algebras in particular, one has that of *correspondence.*

Two systems S_1 and S_2 are said to be in correspondence when there exists a rule which associates to one or more elements of S_1 one or more elements of S_2. The correspondence may be many-one, one-many, many-many, or one-one. This last is said to exist if for every element of S_1 there is associated one and only one element of S_2, and conversely.

Two systems for which a one-one correspondence exists which *preserves all relations* are said to be *isomorphic.* Since the totality of relations in a system is defined to be its structure, one has the very important fact that *isomorphism preserves structure,* or again, that two systems which are isomorphic have the "same" structure. This

*There is a tendency to speak of point-sets nowadays to eliminate confusion with other classes.

last may be taken to define the *equivalence* of two systems.

In many situations, there exists a one-one correspondence which preserves only certain relations and operations, but not all of them. It is convenient to speak then of a *partial isomorphism,* or again of an "isomorphism *in respect*" to these relations and operations. That portion of the structure which is thus preserved can be referred to as a *substructure.*

With the introduction of these notions we shall now turn our attention to several important systems.

2. The Algebra of Number Systems.

Let $(a, b, c, \cdots)$ denote any *integer.* Then we may define for these, two operations, called *addition,* $+$, and *multiplication,* $\times$, such that the following rules are obeyed:

$$\left.\begin{array}{ll}(1) & a + b = b + a\,. \\ (2) & ab = ba\,.\end{array}\right\} \text{(Commutative Law)}$$

$$\left.\begin{array}{ll}(3) & a + (b + c) = (a + b) + c\,. \\ (4) & a(bc) = (ab)c\,.\end{array}\right\} \text{(Associative Law)}$$

$$(5) \quad a(b + c) = ab + ac. \quad \text{(Distributive Law)}$$

Here we have an example of an abstract algebra. It is applicable to the specific integers $1, 2, 3$, etc., $\cdots$, being isomorphic with this system. But clearly, the elements $(a, b, c, \cdots)$ can be made to denote any other system of entities so long as there is isomorphism. That not all systems can be represented by the above can be made clear by an example which Werkmeister (12) gives. For instance, considering addition and the commutative law alone, it is a fact of chemistry that one can add concentrated sulphuric acid to water to get a dilute acid solution, but adding water to concentrated sulphuric acid more likely than not will result in an explosion. Thus the commutative law does not hold in this case. The reader can easily think of other examples of this sort.

The above formulation is just the beginning of the algebra of numbers and really holds only for the integers. It is, however, a relatively easy matter to introduce additional notions into the algebra in order to arrive at one which is isomorphic with the algebra of the real numbers. Such concepts are those of inequality, subtraction, zero, negative and positive integers, fractions, division, powers, and roots. Provided division by zero is forbidden in the resulting algebra, the result is a *closed* system, that is, *one such that all operations per-*

formed within it always produce one of its elements. Rather than expand upon this material, we refer the reader to the very readable account which has been given by Werkmeister (12).

3. Mathematical Representations.

Consider any two isomorphic systems, A and B. These are equivalent in the sense that they have identical structures. By virtue of this, anything which holds true in one holds true in the other, provided the necessary correspondence is established between elements, relations, and operations. In particular, suppose one system, say A, is a physical system,* and the other, B, is some arbitrary algebra derived by a mathematician for his amusement. Should these be isomorphic, that is, should a one-one correspondence exist between the elements of the algebra B and those of the physical system such that all relations are preserved, then the algebra would be a symbolic replica of the physical system. Any relation existing in the physical system will have an image in the algebra. And conversely, for every relation which exists in the latter, one should find a corresponding physical relation. The outcome of such a state of affairs is that we are enabled to speak about the physical system at a purely symbolic level, and even carry out at this level investigations of the properties of the system. We may say, generally speaking, that the algebra constitutes a *symbolic representation* of the physical system, or again that the latter has a representation in the algebra.

Again, speaking in general terms, three situations may arise. The physical system A is isomorphic with either the total algebra B, or only with a subalgebra of this latter. In both instances we shall say that A has a *complete* representation in B, but in addition in the first instance we shall speak of an *equivalent* representation. If on the other hand there is only partial isomorphism with respect to relations in the physical system A, the representation will be said to be *partial.*† Clearly, if A has a complete but not equivalent representation in B, then B can have only a partial representation in A, although it is not usual to speak of algebras as having representations in physical systems.

*The notion of systems as introduced here has been quite general. The elements can be physical entities, events, other systems, and so on. If the elements are physical in nature, then the relations and operations must also be physical. It is here that the great value of being able to establish an isomorphism between physical systems and abstract systems makes itself evident.

†A partial isomorphism in respect to B will still be considered under the heading of complete representation. Partial representation as defined above comes under the heading of *incomplete* representations which also includes cases in which only a subsystem of A is isomorphic with B or part of B.

Of special interest is the situation in which the algebra of numbers constitutes a representation for a physical system. Inasmuch as the remainder of this paper will be largely devoted to just this matter, no more will be said about it here. Instead, a few remarks will be made concerning a particular difficulty which may arise in connection with partial and even complete, but non-equivalent, representations.

Namely, if a physical system does not have an equivalent representation in an algebra, then it is possible to derive relationships in the latter which have no correspondence in the physical system. Even new elements may be defined by the algebra which have no meaning in terms of the second system. In other words, everything which is true in the algebra needs not necessarily hold or have meaning in the physical system. This is just the type of situation which occurs in the domain of applied mathematics when extraneous solutions to equations are obtained. Very often these turn out to be negative or imaginary and quite clearly correspond to nothing in the physical world. While such instances have often led to the discovery of new elements and principles, as in the case of the Dirac electron, it is clear that they may signify nothing more than the fact that an incomplete representation exists.

4. *The Classical Theory of Measurements.*

The present section will be devoted to a brief review of the logic of measurements as expounded by such investigators as Campbell (3), Bergmann and Spence (2), and Reese (8). This material will be familiar to most readers.

In general, it may be said that the basic behavioral process by which scientific data are collected is that of *observation*. Generally speaking, observations can be classified as being *qualitative* or *quantitative*.

A *measurement* is an *operation* as well as an observation performed upon the physical world by an observer and by means of which a certain class of signs are assigned to represent *properties* of physical objects* and events according to certain rules. Such signs are called *numerals*. Any measurable property is called a *magnitude* and the ordered set of all possible numerals which can be assigned to such a property is called a *scale*.†

*Although we will refer here specifically to physical systems, the theory is quite general and applies equally well to biological, psychological, and other kinds of material.

†More extensive discussions of these notions will be found in the works of Werkmeister (12) and particularly of Russell (9).

It is possible to distinguish between two classes of properties, those which are dichotomous or two-valued, and those which are non-dichotomous. To the first category belong such properties as male-female, to-the-right-of, to-the-left-of, and so on. To the second, belong such properties as length, weight, and so on. Non-dichotomous properties of and relations between physical objects and events are said to constitute, or rather to belong to, *physical dimensions.*

There are a number of equivalent ways in which the theory or logic of measurements can be formulated. The following appears to be best suited for the present discussion.

Let X and Y be any two physical events or objects having a property which defines a dimension D. Let $>$ denote the statement "X bears a certain relation $>$ to Y," and let $|>$ denote "X does not bear the relation $>$ to Y." Finally, let $>$ satisfy the following three criteria:

(6) If $X > Y$, then $Y \mid> X$. (Anti-symmetric law)
(7) If $X > Y$ and $Y > Z$, then $X > Z$. (Transitive law)
(8) $X \mid> X$. (Irreflexive law)*

Further, let us define a relation $=$ in terms of $>$ such that:

(9) $X = Y$ if and only if $X \mid> Y$ and $Y \mid> X$.†
(10) If $X = Y$, then $X > Z$ implies $Y > Z$, and $X \mid> Z$ implies $Y \mid> Z$.
(11) From (8), it follows that $X = X$.

Reference to our earlier discussion of ordered systems in Section 1 will show that the relation $>$ as defined above is one of simple ordering. Thus, in respect to the operation $>$, physical systems constitute simply ordered systems. It is for this reason that $>$ is known in the theory of measurements as the relation of physical ordering.

Any physical dimension for which axioms 6 through 11 are satisfied is said to be *intensive.* It is possible to assign numerals‡ $N(X)$, $N(Y)$, and so on, in a non-unique manner to each object of an intensive dimension, such that two relations $>$ and $=$ exist satisfying,

I. $N(X) > N(Y)$ if and only if $X > Y$.
II. $N(X) = N(Y)$ if and only if $X = Y$.

*I have never found this third axiom stated in the literature. Yet it seems to be necessary if measurements are to be unambiguous.

†This is to say that one and only one of the three relations $X > Y$, $Y > X$, or $X = Y$ holds at any given instant of time.

‡Not to be confused with "numbers," although the latter can be used as numerals. In this case their properties as numerals are quite independent of those they possess by virtue of being numbers.

III. $N(X) = N(X)$.*

The resulting ordered set of numerals is called an *ordinal scale.* From the above definition, it is seen that *an ordinal scale is isomorphic with simply ordered physical systems,* when these are described by measurements.

Given a dimension D which satisfies axioms 9 and 10, it will be said to be *extensive* if it is possible to perform a physical operation $+$ on any two members of the dimension such that the result is always another member of the dimension, and such that the following axioms are satisfied:

12. If $X = Y$ and $V = W$, then $X + V = Y + W$.
13. $X + Y = Y + X$.
14. $X + (Y + Z) = (X + Y) + Z$.
15. $X + Y > X$ and $X + Y > Y$.

This new operation, $+$, will be called *addition.*

As previously, numerals can be assigned with an operation $+$ to the properties of an extensive dimension in such a way that,

IV. $N(X + Y) = N(X) + N(Y)$.

The resulting scale is called an *additive scale.*

This much constitutes the essence of what I have called the classical theory of measurements.

5. The Extended Theory of Measurements.

Now arises an interesting situation. Ordinal and additive scales have only partial isomorphism with the algebra of numbers. In fact it is a very restricted isomorphism. This will be seen by comparing the two algebras which have just been presented with that which has been given for the integers. We can, if we desire, use the number system as representation for measurements on ordinal and additive scales, provided one restricts one's self to using only the relations of "greater than," "lesser than," "equal to ," and the operation of "addition."† To subtract measurements from one another at this stage is not allowable, and even less to multiply measurements together— simply because these concepts have not been defined for measurements, although they are well defined for numbers.

At the same time, it becomes clear why numbers may be chosen for numerals, and why, in respect to addition, equality, and inequal-

*See first footnote on preceding page.

†Of course, in the case of ordinal scales, one must also exclude addition.

ity, they have the same properties as numerals that they have as numbers. They are isomorphic in respect to these relations and operations.

Now, aside from the fact that such a restricted isomorphism cannot lead to great consequences, there is also the fact that in practice one finds everyone happily multiplying together, dividing, and otherwise treating measurements as if they were equivalently isomorphic with the most general number systems. Yet, as just shown, this is certainly not true of either intensive or extensive magnitudes.

Campbell and those who have followed in his footsteps have attempted to resolve this problem partly by introducing the notion of derived measures. At this point, unfortunately, these investigators abandon the axiomatic approach with which they started out, and proceed to ignore the very restrictions they had imposed upon measures by setting up their various scales. The net result is something which experientially speaking makes some sense, but which logically and mathematically does not.

Now as it happens, physical scientists resolved the problem in a very ingenious way many years before any one ever tried to write about the logic of physical measurements. They introduced the notion of *dimensional units* and the principle of *dimensional homogeneity*. Accordingly, a number by itself is not sufficient to specify a physical quantity. Its value must be determined by comparing the sample under consideration with a known amount of the same quantity. This constitutes the process of *measuring*.* The quantity used as reference is called the *unit*, and the result of any measurement is a statement of how many times the sample has been found to contain the reference quantity. Thus, if N be a numerical quantity, and U be a unit, a physical quantity, Q, will be represented as,

$$Q = N \cdot U$$

that is, a product of a numerical value and a unit. In consequence, Q is more than a number.

Now, given any physical quantity, there is an infinity of possible units one might use. For instance, take "weight." One can use *"pound," "gram," "drachm,"* and so on. In fact, any arbitrary standard will do provided it possesses in common with all these standards the unique property of being a distinct weight. It is convenient to make use of a general unit symbol, $[U]$, to denote *any* unit of weight. More generally then, we shall write

*It must be understood that we are now speaking in terms belonging to the language of the physical scientist, or more specifically in the language of dimensional analysis.

$$Q = N \cdot [Q]$$

and $[Q]$ will be called the *dimension* of the physical quantity Q, being the expression of a general unit associated with Q.*

The principle of *dimensional homogeneity* simply states that given any equation involving physical quantities, *all terms of the equation,* which are of course physical quantities, *must have the same dimension* if the equation is to have physical sense.

This is a very important point, for without this principle, physical equations could not be treated as mathematical or numerical equations. For as will be seen from the above, dimensions or units have the properties of factors in the writing of physical quantities. Since the principle of dimensional homogeneity requires a common dimension throughout any equation, the latter may be cancelled out in each term, leaving only a numerical quantity to be dealt with.

To go further into the topic of dimensional analysis is beyond the scope of this paper, but the reader will be well-rewarded in looking up a more detailed treatment of this topic, such as has been presented by Bridgman (1), and by Eshbach (4). Now we must return to the axiomatic treatment of the problem.

If the reader will refer to axioms 6 through 11, he will note that by their very nature they require the elements for which the relations $>$, and $=$, and the operation $+$ hold to be of the same dimension. The principle of dimensional homogeneity is therefore inherent in the definition of ordinal and additive scales. But this is just about as far as the resemblance between the two approaches goes. For while in the system developed by dimensional analysis any physical equation is reducible to a numerical equation, and hence physical systems have general mathematical representation, that is not possible with the two scales developed by axiomatic methods.

I wish to propose at this point that physical measurements or magnitudes lie on a scale which is neither the ordinal nor the additive type of scale, nor a "derived" scale.

For lack of a better descriptive term, let us say that a dimension D satisfying all of the criteria of an extensive dimension is *supra-extensive* if it is possible to perform a physical operation $\cdot$ on any two members of the dimension such that the result is always a member *of another* supra-extensive dimension, and such that the following axioms are satisfied:

*Those readers who are familiar with Cantor's definition of transfinite numbers will see here a notion defined in a manner very much like the definition of cardinal numbers.

16. If $X = Y$ and $V = W$, then $X \cdot V = Y \cdot W$.
17. $X \cdot Y = Y \cdot X$.
18. $X \cdot (Y \cdot Z) = (X \cdot Y) \cdot Z$.
19. $X \cdot (Y + Z) = X \cdot Y + X \cdot Z$, if and only if Y and Z are of the *same* dimension.

The new operation $\cdot$ will be called *multiplication*. As previously, one can assign numerals with the operation $\cdot$ to the properties of a supra-extensive dimension in such a way that,

V. $N(X \cdot Y) = N(X) \cdot N(Y)$.

The resulting scale will be called a *multiplicative* scale. It has of course all the properties of additive and ordinal scales too.

It is important here not to confuse "multiplication by an integer" with "multiplication" as defined for supra-extensive magnitudes. The former is, strictly speaking, a property of additive scales. More specifically, if a quantity A be added n times to itself (iterative addition), we may denote this by the "product" nA, where n is an integer, and where nA stands for the sum $(A + A + A + \cdots + A)$ of n terms. This is not the same thing as the "product" $A \cdot A$ defined earlier. The difference between the two kinds of products is the same here as it is, for instance, in the case of vectors,* where the product $n\bar{A}$ of a vector $\bar{A}$ by a scalar quantity n is to be distinguished from the vector product $\bar{A} \times \bar{B}$, or even the scalar product $\bar{A} \cdot \bar{B}$, of two vectors $\bar{A}$ and $\bar{B}$. The former is a vector of magnitude n times that of A and of similar orientation, while in general, the vector product of $\bar{A}$ and $\bar{B}$ is another vector at right angle to both of these and with magnitude equal to the product of their magnitudes times the sine of the angle they make with each other. In particular, $\bar{A} \times \bar{A} = 0$, and generally, $\bar{A} \times \bar{B}$ does not equal $\bar{B} \times \bar{A}$. The properties of scalar products are still different. All of this is entirely consistent with the fact that vectors are not single numbers, but are really pairs of the same, that is, dyads, and as such have properties not common to cardinal numbers. In particular, vectors exist on a multiplicative scale not by virtue of the existence of products of the form $n\bar{A}$, but because of such products as $\bar{A} \times \bar{B}$.

Along the same line of thought as the above, one should not confuse the notion of multiplication by a *constant* with multiplication by

*It may be of interest to note here that fairly recently a new approach to dimensional analysis has been developed in terms of vectors in an affine space.

a *numerical coefficient.* In closed systems, members of the system alone can be employed, and multiplication is as defined for the system. On the other hand, introduction of a numerical coefficient is equivalent to "multiplication by an integer" and is not allowable in a closed system unless some new operation such as the scalar product for vectors is defined.

It may be well to add one last remark concerning the establishment of multiplicative scales. Namely, inasmuch as multiplication as here defined always produces a supra-extensive dimension different from the dimensions of the factors in the product, it necessarily follows that this operation creates a relation between two or more different supra-extensive dimensions. It bridges the gap, so to speak, between different supra-extensive dimensions. This is a property which is unique and not found in the cases of intensive and extensive magnitudes. In turn, this means that corresponding relations must exist between the phenomena being measured. In practice, it is the very demonstration in most instances of such relations between phenomena which establish magnitudes on a multiplicative scale. That is, it is only by working with and causing entities belonging to different supra-extensive dimensions to interact that one can formulate the multiplicative character of the corresponding magnitudes. For instance, voltage considered alone has only additive properties. Similarly for current. It is only when we consider the two together in term of power that their multiplicative properties become apparent.

There remains now very little more to be done in extending the theory of measurements. "Subtraction," "zero," and "multiplication by an integer" could have been introduced earlier by standard axiomatic methods in terms of the properties of additive scales before the notion of multiplicative scales was formulated. On the other hand, the notions of "division," "powers," "roots," "logarithms," and many others had to wait until the new scales had been defined. It would be instructive to develop the entire system thus outlined. As it is, space for this is lacking, and it may be said only that this development follows exactly the same steps as does that of the algebra of number systems. It is, however, not hard to see, even without doing this, that *multiplicative scales lead to an algebra which is isomorphic with the arithmetic (or algebra) of the real numbers.* Of the three scales which have been discussed, *it is the only one* which can do this. It is therefore the only one which can lead to a mathematical representation of physical phenomena.

6. *Some Consequences for Psychology.*

I began this inquiry by asking why it is that mathematical representations have been so successful in the physical sciences and yet have so completely failed us in psychology, even in those areas where measurements are possible. I believe part of the answer lies in the previous pages.

It is my hope that it has been shown in a sufficiently clear manner that *the possibility of adequate mathematical representation for any system, physical or non-physical, depends upon the possibility of establishing a system of measurements which is isomorphic with the number system and other mathematical systems.* That is, it must be possible to replace the physical system by the mathematical system. In turn, this means that the two systems have the same structure or are equivalent. Even with partial isomorphism the representation may still be quite satisfactory provided a sufficiently large portion of the structures involved are preserved. On the other hand, nothing can ever be as satisfactory as equivalent representation. The physical sciences have come close to achieving this ideal through their use of dimensional units and analysis. In the case of psychology, the situation is pretty much the opposite. Psychological measurements are largely made in terms of additive scales, and often using only ordinal scales. Consequently, the correspondence which may be established between psychological magnitudes and numbers is very limited. As can then be expected, mathematical representation of psychological phenomena is quasi-impossible, being far too restricted. This is not to say that improvements in the right direction have not been made in recent years. For instance, the use of dimensional units has been introduced for some subjective magnitudes. Thus we have a *sone* scale, a *mel* scale, and a *veg* scale, to mention only a few. And again, in learning theory, we have seen the introduction of various units, such as the *hab*. Thus far, however, most of the data of psychology, if not all, remain within the bounds of additive and ordinal scales. Even Stevens' (10) promising "ratio scale" turns out to be, as Reese (8) has previously observed, nothing more than an additive scale. This follows simply from the observation made earlier concerning the difference between "multiplication by an integer," and "multiplication" as understood for supra-extensive magnitudes. Or to say this another way, to state that a subjective magnitude is, for instance, half of another, is to say nothing more than the second is twice the first, this being expressible entirely in terms of addition. In addition to limitations imposed by the scales used, many psychological vari-

ables are "quantized," that is, they can take on only certain values. Thus, for instance, the number of trials made in a learning experiment cannot be fractional, or irrational, or negative, but must be a positive integral value. This fact imposes some rather serious restrictions upon the means available for representation.

Failure to recognize the fact that psychological magnitudes rarely have all of the properties of numbers has led to the formulation of mathematical equations which have a superficial appearance of validity, if not of mathematical sanction, but which are inherently unsound and certainly misleading.

A few examples of what is meant might be given. For instance, consider the well-known Fechner relation,

$$S = k \log R \,,$$

S being the subjective magnitude of the stimulus (sensation), and R the physical magnitude of the stimulus (stimulus intensity). One may, first of all, note a certain inherent ambiguity in the relation. For, $\log R$ is a pure number, as can be shown by dimensional analysis, regardless of the dimensions of R. If k also be assumed dimensionless, then the Principle of Dimensional Homogeneity requires the same be true of S. On the other hand, if we assume S to have some arbitrary dimension, then by the same principle, k must have the inverse dimension of S. Unfortunately, we have no way of deciding which of the alternatives is the correct one. Hence the essential ambiguity in this law. However, a much more serious defect in the relation lies in the fact that in general, S is *at best* only an additive quantity, while $k \log R$ has *all* of the properties of real numbers, being one itself. The equality relation which holds between the two members of the relation is therefore a contradiction of these facts. Actually, as written, the Fechner Law states that a quantity with only *some* of the properties of real numbers is the same as a real number. To make this example even more specific, we might consider Sanford's lifted weight experiment as presented by Guilford (5). Briefly, a subject was requested to lift various weights and to make five groups of these in such a way that the differences in the weights between neighboring groups would appear to be equal to him. In other words, the method of equal-appearing intervals was employed. This done, "subjective" values 1, 2, 3, 4, and 5 were assigned to the groups in order of increasing weights. The actual mean weights for each group were taken as the corresponding "physical" weights. These turned out to be: 6.52, 10.88, 11.87, 43.96, and 79.52 gm.

Plotted one against the other, these two sets of magnitudes yield a logarithmic type of curve. Or again, plotting subjective values against the logarithm of the physical weights produces a straight line. Thus, the Fechner law appears to be substantiated in this instance.

But, inasmuch as the method of equal-appearing intervals does not produce anything better than additive quantities, and since $k \log R$ in this case is a real number, the equality between S and $k \log R$ is only partly true. Furthermore, the values taken by S are quantized. But without additional qualifying restrictions upon the relation which appears implied, one is led to the prediction of the existence of an innumerable quantity of *meaningless* values of S, namely, irrational values.

Such a situation is not unique to psychology. As a matter of fact, something of this sort has occurred in mathematics itself. For centuries, mathematicians tried to trisect exactly, by means of compass and ruler only, any angle whatsoever. While it is perfectly feasible and meaningful to calculate one third of an arbitrary angle, it is quite another matter to go through the physical operation of doing this in an exact manner with the above mentioned tools. As it eventually turned out, the problem is insoluble, because as was finally shown, dividing an angle by three does not admit of any representation in terms of compass and ruler. Said another way, the operations and relations one may obtain with a compass and ruler give rise to a structure which is not the same as that existing for numbers.

Fechner's Law is only one of the forms the general function $S = f(R)$ may take. For instance, Stevens and Harper (**11**), using a "ratio scale," find that "subjective weight" and "physical weight" are related according to

$$\log S = 14.58 \log (1 + \log R) - 6.94 .$$

This is a far cry from Fechner's Law. Yet much of what has already been said concerning it applies here. Furthermore, the entire derivation of the above relation appears questionable. For, to do so, the investigators began by plotting $\log S$ against $\log R$. But, as already indicated, S measured on a ratio scale *only has additive properties*, while taking the logarithm of S *presupposes or implies that* S *has multiplicative properties.** Stated somewhat differently, one can-

*This follows from the fact that $\log x = \frac{x-1}{x} + \frac{1}{2}\left(\frac{x-1}{x}\right)^2 + \frac{1}{3}\left(\frac{x-1}{x}\right)^3 + \cdots$

not take the logarithm of S because in this instance it does not exist. In any case, it is not permissible to make the above plot.

Psychophysical laws are not the only ones to come under the above criticism. For instance, Hull (6) describes an experiment in learning in which the variables are "number of reactions to produce extinction" (n), and "number of reinforcement repetitions" (N). After plotting values of these two variables, he deduces that they are related according to the law

$$n = M(1 - 10^{-kN}) + C.$$

Inasmuch as actually n and N can both only have integral values, while, on the contrary, the above places no such restrictions upon the values n and N may take, it hardly can be said that this relation is a mathematical representation of the true state of affairs.

Similarly, elsewhere, Hull (6) plots the "number of j.n.d.'s distant from the point of reinforcement" (d), against the "amplitude of galvanic skin response" (A). This plot, he claims, can be represented by the relation

$$A = 18.3 - 6(1 - 10^{-.0135d}).$$

Again, as stated here, the relation does not impose any restrictions upon the values d may take. As a matter of fact, it is mathematically allowable and entirely feasible to solve the equation for d. Mathematically speaking, d can be a fractional and even an irrational quantity and satisfy the above. Yet, to speak of fractional j.n.d.'s and even more so of irrational amounts of j.n.d.'s is rather meaningless. In fact, it is a self-contradiction!* Actually, however, even if the necessary restrictions could be incorporated in the above, there remains the fact that the values of A lie on a multiplicative scale, while those of d, being j.n.d.'s are not of this type. Consequently, the above relation once more equates quantities which are fundamentally different in respect to mathematical structure.

One last example of the kind of difficulties which may arise from faulty representations may be given. Some years back, considerable effort was directed at establishing a relationship between brain mass, or cortical area, and intelligence. Possibly less energy would have been expended along this line if the inherent fallacy of equating intelligence to a function of brain mass or of cortical area had been recognized. For suppose, to make the example simple, that one postulated and even found a trend that I.Q $= kM$ (M being brain mass).

*Since the j.n.d. is by definition a "just noticeable difference," a fraction of j.n.d. would certainly not be noticeable and therefore could not be determined, or rather defined, empirically.

Suppose, too, that two individuals were found with brain masses M_1 and M_2 such that $M_1 = 2M_2$. Then it would follow that $\text{I.Q.}_1 = kM_1 = 2kM_2 = 2 \cdot \text{I.Q.}_2$. Since, however, I.Q.'s are not additive quantities, such a result is obviously a contradiction of the true state of affairs, as actual measurements would indeed show. This, one could have predicted from the outset, since mass and area are multiplicative quantities, while I.Q.'s are not.

One could go on in this manner *ad infinitum*. In brief, the chief difficulty appears to boil down to the fact that few, if any, experimental and theoretical data in psychology are ever given an appropriate mathematical representation. In nearly every instance, the relations which are developed assign to the psychological quantities properties which they do not and cannot have.

But, many psychologists will remark, we can plot our data, and certainly we do obtain geometrical figures to which mathematical formulas correspond. How can all this be? The answer to this is that plotting is no more permissible than writing down mathematical equations in the instances already cited. The reason for this is a perfect example of what may be done when structures are equivalent. Namely, there is an isomorphism between the points in a plane and pairs of numbers. What is incorrect to do in respect to one is also incorrect in respect to the other. If this were not so, the isomorphism could not be true. There is ample evidence, of course, that such is not the case. It is entirely permissible to use mathematical equations to denote symbolically a particular *trend*. One can go a step further and manipulate the equations, *provided this is done within the limitations* imposed by the properties of the magnitudes which are involved. It is very tempting when one has numbers for data to generalize, replace these by variables, and write equations. It is maybe even more tempting to go on and solve for various variables in terms of the others, to take derivatives, integrate, and do numerous other things. The trouble unfortunately is that, as I have pointed out, *"numbers" when used as "numerals" do not always have all of the properties of "numbers."* This is a fact which the majority of psychologists appear either to forget or to be completely unaware of it. A simple solution to this state of affairs would be to employ numerals which can be distinguished from numbers as such and to set down in every case the rules, operations, and so on which are applicable. If this were done, for instance, for a non-multiplicative quantity A, one would never be led to plot $\log N(A)$* against some other variable since this par-

*In accord with the notation of page 392, $N(A)$ denotes any numerical associated with magnitude A.

ticular function is not definable and hence is meaningless in the case of the magnitude A . But if one uses a number like "5" as a numeral in the same instance, it is much too easy to overlook the fact that "log 5" is still not definable.

Yet, it should be clear that while paying strict attention to the above considerations would be a considerable improvement in psychology, it is doubtful that it would allow this discipline to attain equal footing with the physical sciences in respect to the use of mathematical representation. For, as I have tried to emphasize, the structure of psychological measurements as they exist today admits of *too limited* a mathematical representation. Seeking and developing special branches of mathematics which would have suitable structures would lead to representation which would be too limited in number and which would tend to be much too broad generalities. A typical instance of this sort of situation I have in mind is Lewin's (7) attempt to apply topology to psychology.

By far a more rational and promising approach would appear to lie in the direction of (a) developing multiplicative scales in psychological measurements, and (b) in redefining basic notions in terms of magnitudes which are susceptible to being measured on such types of scales. In other words, our best hope appears to lie in ending our efforts to force obviously unsuitable data into existing mathematical structures and trying instead to develop methods for obtaining more suitable material, if it exists. In particular, in accordance with a remark made earlier in connection with multiplicative scales, special effort should be made toward formulating relationships which have empirical correlates between psychological magnitudes. The more, the better. Quantities like "subjective weight (or force)," and "subjective length" have little value by themselves. In the final analysis, relating these to physical quantities, as is done in Fechner's Law, does not appear to me to be particularly significant in respect to understanding how the mind works. In any event, as shown previously, its use has numerous pitfalls. On the other hand, I believe that defining a concept such as "subjective work" and identifying it with the "product" of "subjective force" and "subjective length" is far more likely to lead to significant results. For if at a future date it becomes possible to associate with it actual measurements, then an important step will have been taken toward creating multiplicative psychological magnitudes, and the first step toward *adequate mathematical representation* in psychology will have been made.

REFERENCES

1. Bridgman, P. W. Dimensional analysis. New Haven: Yale Univ. Press, 1922.
2. Bergmann, G., and Spence, K. W. The logic of psychophysical measurement. *Psychol. Rev.*, 1944, 51, 1-24.
3. Campbell, N. R. An account of the principles of measurements and calculations. London: Longmans, Green, 1928.
4. Eshbach, O. W. Handbook of engineering fundamentals, Vol. I. New York: John Wiley and Sons, Inc., 1936.
5. Guilford, J. P. Psychometric methods. New York: McGraw-Hill Book Co., 1936.
6. Hull, C. L. Principles of behavior. New York: Appleton-Century Co., 1943.
7. Lewin, K. Principles of topological psychology. New York: McGraw-Hill Book Co., 1936.
8. Reese, T. W. Application of the theory of physical measurement to the measurement of psychological magnitudes with three experimental examples. *Psychol. Monogr.*, 1943, 55, No. 3.
9. Russell, B. Principles of mathematics. New York: W. W. Norton and Co., 1943.
10. Stevens, S. S. On the theory of scales of measurement. *Science*, 1946, 103, 677-680.
11. Stevens, S. S., and Harper, R. S. A psychological scale of weight and a formula for its derivation. *Amer. J. Psychol.*, 1948, 61, 343-351.
12. Werkmeister, W. H. The basis and structure of knowledge. New York: Harper and Brothers, 1948.

Manuscript received 1/8/51
Revised manuscript received 4/16/51

PSYCHOMETRIKA—VOL. 16, NO. 4
DECEMBER, 1951

ESTIMATION OF THE RELIABILITY OF RATINGS*

ROBERT L. EBEL
THE STATE UNIVERSITY OF IOWA

A procedure for estimating the reliability of sets of ratings, test scores, or other measures is described and illustrated. This procedure, based upon analysis of variance, may be applied both in the special case where a complete set of ratings from each of k sources is available for each of n subjects, and in the general case where $k_1, k_2, \cdots, k_n$ ratings are available for each of the n subjects. It may be used to obtain either a unique estimate or a confidence interval for the reliability of either the component ratings or their averages. The relations of this procedure to others intended to serve the same purpose are considered algebraically and illustrated numerically.

The Problem

The process of estimating test reliability by correlating two sets of scores is well known. The two sets of scores are usually obtained from two equivalent forms, split halves, or two administrations of the test. But when one is dealing with measures other than test scores, such as performance ratings, it frequently happens that more than two parallel sets are available. For example, in one study which concerned us recently, nineteen English instructors graded each of five themes. We desired a measure of their agreement with each other, both before they had received special training in theme rating and again after that training. A complete table of ratings was available in this study, since each instructor rated each of the themes.

In other similar studies, however, the available sets of ratings are sometimes incomplete. An example of this is provided by a study in which eight physics professors rated the research potentialities of twenty-two graduate students. We wished to measure the agreement among the raters in order to establish the reliability of the average ratings as a criterion for validating a selection test. Each professor was asked to rate only those students whose work he knew well. Hence the table of ratings in this study was incomplete.

Problems similar to those just mentioned arise frequently in educational and psychological research. Several formulas have been pro-

*The writer wishes to acknowledge the helpful comments and suggestions of Professors E. E. Cureton, Harold Gulliksen, and E. F. Lindquist.

posed to deal with them, but there has been no general agreement on a best method. Peters and Van Voorhis (8) present a formula for average intercorrelation (No. 118) which is appropriate in certain situations. While this formula is derived on the basis of a complete table of ratings, further derivation leads to another formula (No. 119) which may be used where ratings are incomplete. Clark (1) has reported a study in which such a formula was applied to data which did not provide a complete table of ratings.

Fisher's work on intraclass correlation (3) has led to a formula based upon the analysis of variance. This formula is presented in convenient form by Snedecor (9). It is applicable to either complete or incomplete sets of ratings. Horst (5) developed a generalized formula for the reliability of measures which is also applicable to either complete or incomplete sets of ratings. Horst's formula yields the reliability of *average* ratings. However, the Spearman-Brown transformation may be used to obtain the reliability of individual ratings.

At this point a question is likely to arise. Is it better to estimate the reliability of individual ratings or the reliability of average ratings? If decisions are based upon average ratings, it of course follows that the reliability with which one should be concerned is the reliability of those averages. However, if the raters ordinarily work individually, and if multiple scores for the same theme or student are only available in experimental situations, then the reliability of individual ratings is the appropriate measure. Since the reliability of average ratings is determined completely by the reliability of the component ratings, and by the number of components, it is always possible to determine the reliability of individual ratings, or of averages, no matter which value a formula gives initially. Formulas using both approaches will be presented in the following section.

A somewhat different approach to the problem of rater agreement has been suggested by Gulliksen (4). This approach, based on the Wilks-Votaw tests for compound symmetry, does not yield a quantitative estimate of the *degree of agreement* between ratings, but provides instead a statistical criterion upon which to base a categorical statement that the raters do or do not agree. If the distributions of scores from different sources are similar enough so that one can not reject at the five per cent level of confidence the hypothesis that the sets of ratings are random samples from the same population, Gulliksen recommends that the sources of ratings be regarded as parallel (or in agreement). If this hypothesis can be rejected at the one per cent level of confidence, he recommends that the sources

be regarded as not parallel. Gulliksen's suggestion has another application in the study of rater agreement. If one has *a priori* reason to believe that different "schools of thought" may exist among the raters, it is possible to use the Wilks-Votaw tests to check this hypothesis. One should not, however, test the hypothesis on the same set of data that suggested it.

The Intraclass Correlation Formula

When the formulas of Peters, Snedecor, and Horst for estimating the reliability of ratings were applied to the same sets of data, they yielded some inconsistent coefficients. An analysis of the sources of these inconsistencies has led to the conclusion that the formula for intraclass correlation is the most convenient and generally useful. The derivation of this formula is outlined here, since it has not been widely used in studies of educational and psychological ratings, and since few textbooks on measurement contain any discussion of it.

Suppose we have a sample of k estimates of a trait in each of a sample of n persons. Each estimate may be considered to consist of a true component and an error. The true component is constant in all k estimates for any one person, but varies from person to person. Let A represent the variance of these true components in the population of persons from which we have sampled.

The error component varies from estimate to estimate for the same person, but this variance is assumed to be substantially the same in all sets of ratings for the various persons. Let B represent the variance of these errors in the population of estimates. The total observed variance of the estimates is thus $A + B$.

The reliability of the estimates is defined as that portion of the observed variance which is true variance, or

$$r = \frac{A}{A + B}. \tag{1}$$

Suppose we have analyzed the variance of the foregoing sample of estimates to obtain a mean square for error (M) and a mean square for persons (M_p). The mean square for error is a direct estimate of B, the variance of the population of errors of estimate, or

$$M = B. \tag{2}$$

The mean square for persons, however, is not a direct estimate of A. Rather, it represents k times the variance of the means of the estimates for each of the n persons. The variance of these n means is not

attributable to A alone, but also includes an error component attributable to B. Each mean consists of a true component, drawn from a population with variance A, and an error component, which is the mean of k errors drawn from a population with variance B. Hence the variance of the means is $A + B/k$. The mean square for persons is k times the variance of the means. Hence

$$M_{\bar{x}} = kA + B. \tag{3}$$

Solving equations (2) and (3) for A and B, and substituting these values in formula (1), we obtain the formula given by Snedecor,

$$r_1 = \frac{M_{\bar{x}} - M}{M_{\bar{x}} + (k-1)M}. \tag{4}$$

This is the formula for the reliability of individual ratings. Cureton (2) suggests the following parallel derivation of a formula for the reliability of average ratings.
For the average scores, the variance ratio analogous to (1) is

$$r_k = \frac{A}{A + \bar{B}}, \tag{a}$$

where $\bar{B}$ is the error variance of the person means. The estimate of A is still given by (3) and (2); i.e.,

$$A = \frac{M_{\bar{x}} - M}{k}, \tag{b}$$

and the estimate of $\bar{B}$ is given by the usual formula for the error variance of a mean,

$$\bar{B} = \frac{M}{k}. \tag{c}$$

Substituting from (b) and (c) in (a), we obtain at once,

$$r_k = \frac{M_{\bar{x}} - M}{M_{\bar{x}}}. \tag{d}$$

It is worth noting that formula (d) may also be derived by applying the Spearman-Brown formula to formula (4) given above.

TABLE 1
Analysis of Ratings for Problem 1—Complete Sets

	Rater 1	Rater 2	Sum	Sum^2
Pupil 1	3	1	4	16
Pupil 2	1	3	4	16
Pupil 3	5	4	9	81
Pupil 4	4	5	9	81
Sum	13	13	26	194
Sum^2	169	169	338	

Sum of squared ratings		=	102
Product of sum and mean	$26 \times \frac{26}{8}$	=	84.5
Sum of squares			
For raters	$\frac{338}{4} - 84.5$	=	0.0
For pupils	$\frac{194}{2} - 84.5$	=	12.5
For total	$102 - 84.5$	=	17.5
For error	$17.5 - 12.5 - 0.0$	=	5.0
Mean square			
For pupils	$12.5 \div 3$	=	4.1667
For error	$5.0 \div 3$	=	1.6667
Reliability of ratings	$\frac{4.1667 - 1.6667}{4.1667 + (2-1)\ 1.6667}$	=	.4286
Reliability of average ratings	$\frac{4.1667 - 1.6667}{4.1667}$	=	.6000

Illustrations

Snedecor's formula is applied to a simple problem involving a complete table of ratings in Table 1. Those not familiar with analysis of variance may refer to Table 4 for the formulas used. In this analysis, three components, attributable to pupils, raters, and error, may be separated. Thus it is possible, if desired, to remove the "between-raters" variance from the error term. This overcomes the chief objection of Peters and Van Voorhis to intraclass correlation coefficients, which is that such coefficients are seriously distorted by differences between raters in general level of rating. In Table 1, the "between-raters" variance is zero, so retention or removal gives the same result.

Whether or not it is desirable to remove "between-raters" variance in estimating the reliability of ratings depends upon the way

in which the ratings are ultimately used in grading, classification, or selection. In any case where differences from rater to rater in general level of rating do not lead to corresponding differences in the ultimate grades, classifications, or selections, the "between-raters" variance should be removed from the error term. Specifically, the "between-raters" variance *should be removed* where the final ratings on which decisions are based consist of averages of complete sets of ratings from all observers, or ratings which have been equated from rater to rater such as ranks, Z-scores, etc. Likewise, if comparisons are never made practically, but only experimentally, between ratings of pupils by different raters, the "between-raters" variance should be removed. But if decisions are made in practice by comparing single "raw" scores assigned to different pupils by different raters, or by comparing averages which come from different groups of raters, then the "between-raters" variance should be included as part of the error terms.

TABLE 2

Analysis of Ratings for Problem 5—Incomplete Sets

	Ratings	k	Sum
Pupil 1	8 6 4 4 3	5	25
Pupil 2	6 9 9 4 9 6 5 10 8	9	66
Pupil 3	4 9 10	3	23
	Sums	17	114
Sum of squared ratings			858
Sum of products (pupil sum times pupil mean)			785.3333
Product of sum and mean			764.4706
Sum of squares			
For total	858 — 764.4706	=	93.5294
For pupils	785.3333 — 764.4706	=	20.8627
For error	93.5294 — 20.8627	=	72.6667
Mean square			
For pupils	20.8627 ÷ 2	=	10.4314
For error	72.6667 ÷ 14	=	5.1905
Average value of k			5.1176
Reliability	$\dfrac{10.4314 - 5.1905}{10.4314 + (4.1176)(5.1905)}$	=	.1648

Table 2 illustrates application of this formula to a simple problem in which the table of ratings is incomplete and the sources of ratings are not identified. In this case only two components of the variance, attributable to pupils and error, are separated. Thus any difference in general level of ratings between the various raters is automatically included in the error term.

The application of this reliability formula in Table 2 presents a special problem. The formula requires a value of k, the number of ratings of each person. But in Table 2 this number is not the same for each person. Snedecor (9, p. 234) suggests the following formula for an average k:

$$k_0 = \frac{1}{n-1}\left[\sum k - \frac{\sum k^2}{\sum k}\right]. \tag{5}$$

The average k thus obtained is approximately the harmonic mean of the k's for each pupil.

Estimates of Precision

When the reliability of ratings is estimated on the basis of a sample of products or raters or both, a description of the precision of the reliability estimate obtained is useful in judging the adequacy of the sample or the confidence which can be placed in the obtained estimate. The intraclass formula lends itself readily to such a description. The method used here was suggested by Jackson and Ferguson's description of confidence intervals for their sensitivity coefficient (6).

The first step in obtaining this estimate is to express the variance between products, and the error variance as a single ratio, here designated by F_s. In equation form

$$F_s = \frac{M_{\bar{x}}}{M}. \tag{6}$$

Using this ratio it is possible to transform the intraclass formula as follows:

$$r = \frac{F_s - 1}{(F_s - 1) + k}. \tag{7}$$

Now if the sample values of $M_{\bar{x}}$ and M are considered as having been drawn from separate populations, it is obvious that the ratio (F_p) for the populations might be either greater or less than the ratio observed in the sample. How much greater or less one should believe it to be, at any selected level of confidence, can be read from a Table for F, given the degrees of freedom for the sample values of $M_{\bar{x}}$ and M. To obtain an estimate for the upper limit of the variance ratio between two populations, one must multiply the variance ratio

observed in the samples (F_s) by the maximum ratio expected if the populations had been equal in variance. This value (F_t) may be read in the table, given the number of degrees of freedom in the samples, and the level of confidence desired. In entering the table for F it is important to remember that at the upper limit $M_{\bar{x}}$ is the larger variance, so that the degrees of freedom for $M_{\bar{x}}$ should be located on the marginal headings for the larger variance. Similarly, to obtain the lower limit of (F_p) one must multiply the sample value by the reciprocal of the value given in the table. Remember that at the lower limit, M is the larger variance, so that in this case the degrees of freedom for M should be located on the marginal headings for the larger variance. Substituting these two limiting values for the population variance ratio in formula 6 yields upper and lower limits of the confidence interval for the estimate of reliability obtained from a particular sample.

The application of this procedure to several sample problems is illustrated in Table 3. The data for problems 1 and 3 are taken from

TABLE 3
Confidence Limits (5%) for Reliability Coefficients From Sample Data

Problem	1	3	5
Unique Estimate	.4286	.3429	.1648
k	2.	2.	5.1176
F_s	2.500	2.0435	2.0096
F_t (upper 5% limit)	9.28	9.28	3.74
(lower 5% limit)	9.28	9.28	19.42
F_p (upper 5% limit)	23.125	18.964	7.5159
(lower 5% limit)	.2694	.2202	.1035
r (upper 5% limit)	.92	.90	.56
(lower 5% limit)	—.58	—.64	—.21

Table 5, and the data for problem 5 are taken from Table 2. Because of the smallness of the samples involved, the confidence limits for these reliability coefficients are very wide indeed. Reliability estimates based upon such small samples are little better than blind guesses.

Relationships Among the Formulas

The fact that various formulas have been presented for estimating the reliability of ratings, and that these formulas do not always

yield consistent results when applied to the same set of data was mentioned in the first section of this paper. It is therefore appropriate to consider the relationships among them first analytically and then, by way of illustration, empirically. Consider first the relation between the Pearson product-moment formula and the formula for average intercorrelation given by Peters and Van Voorhis. As usually presented these formulas are

(1) Product moment:

$$r = \frac{\sum xy}{N\,\sigma_x\,\sigma_y}. \tag{8}$$

(2) Average intercorrelation:

$$r = \frac{(\sigma_s^2/\sigma_i^2) - a}{a^2 - a}. \tag{9}$$

We note that the quantity $\frac{\sum xy}{N}$ is the covariance of the scores for each *product* from two *observers*. The symbol ζ will be used to represent this covariance.

It is easy to show (8, p. 196) that the variance of the sums of scores from k observers is

$$\sigma_s^2 = k\,\sigma_i^2 + (k^2 - k)\zeta_{ij}. \tag{10}$$

If this value is substituted for σ_s^2 in formula (9) the result is

$$r = \frac{\zeta_{ij}}{\sigma_i^2}. \tag{11}$$

Since the product-moment formula may be written

$$r = \frac{\zeta}{\sqrt{\sigma_x^2\,\sigma_y^2}}, \tag{12}$$

it is clear that when $k = 2$, (the only case in which the product-moment formula is applicable) and when the table of ratings is complete (which must be true for either formula to apply) the numerators are identical. But since σ_i^2, which represents the variance of scores from either observer, is estimated by taking the *arithmetic mean* of the variances of scores from each observer, σ_i^2 will not equal the geometric mean of these variances, $\sqrt{\sigma_x^2\,\sigma_y^2}$ except in the special case where $\sigma_x^2 = \sigma_y^2$. Hence, differences in coefficients obtained from these two formulas are attributable to the difference in methods of calcu-

lating the average variance within observers.

It is interesting to note in passing that ζ provides a direct estimate of A, the variance of the population of true scores. Recalling the definitions of A and B given earlier in this paper, one may see that

$$\sigma_s^2 = k^2 A + k B \tag{13}$$

TABLE 4

Notation for Analysis of Ratings

(Complete Arrays)

A. Data

n products	k observers								Sums
	0_1	0_2	0_3	. . .	0_j	. . .	0_k		
P_1	x_{11}	x_{12}	x_{13}	. . .	x_{1j}	. . .	x_{1k}		P_1
P_2	x_{21}	x_{22}	x_{23}	. . .	x_{2j}	. . .	x_{2k}		P_2
P_3	x_{31}	x_{32}	x_{33}	. . .	x_{3j}	. . .	x_{3k}		P_3
.	.	.	.		.		.		.
.	.	.	.		.		.		.
.	.	.	.		.		.		.
.	.	.	.		.		.		.
P_i	x_{i1}	x_{i2}	x_{i3}	. . .	x_{ij}	. . .	x_{ik}		P_i
.	.	.	.		.		.		.
.	.	.	.		.		.		.
.	.	.	.		.		.		.
.	.	.	.		.		.		.
P_n	x_{n1}	x_{n2}	x_{n3}	. . .	x_{nj}	. . .	x_{nk}		P_n
Sums	0_1	0_2	0_3	. . .	0_j	. . .	0_k		T

B. Analysis of Variance

Source	Degrees of Freedom	Sums of Squares	Mean Square
Products	$n-1$	$S_p = \frac{\Sigma P^2}{k} - \frac{T^2}{nk}$	$M_p = \frac{S_p}{n-1}$
Observers	$k-1$	$S_0 = \frac{\Sigma 0^2}{n} - \frac{T^2}{nk}$	$M_0 = \frac{S_0}{k-1}$
Error	$(n-1)(k-1)$	$S_e = S_t - S_p - S_0$	$M_e = \frac{S_e}{(n-1)(k-1)}$
Total	$nk-1$	$S_t = \Sigma x^2 - \frac{T^2}{nk}$	

and

$$\sigma_i^2 = A + B\,. \tag{14}$$

Substituting these in (9) and simplifying leads to the expression

$$A = \zeta\,. \tag{15}$$

Consider next the relation between the formulas for average intercorrelation (9) and intraclass correlation (4). The notation used in this and the following comparison is presented in Table 4. Since

$$\sigma_s^2 = \frac{\sum P^2}{n} - \left(\frac{T}{n}\right)^2$$

or

$$n\,\sigma_s^2 = \sum P^2 - \frac{T^2}{n},$$

and since

$$S_p = \frac{\sum P^2}{k} - \frac{T^2}{nk}$$

or

$$k\,S_p = \sum P^2 - \frac{T^2}{n},$$

it follows that

$$\sigma_s^2 = \frac{k\,S_p}{n}. \tag{16}$$

Further, since

$$\sigma_i^2 = \frac{\sum\sum x^2}{nk} - \frac{\sum 0^2}{n^2k}$$

or

$$n\,k\,\sigma_i^2 = \sum\sum x^2 - \frac{\sum 0^2}{n},$$

and since

$$S_e + S_p = S_t - S_0 = \sum\sum x^2 - \frac{\sum 0^2}{n},$$

then

$$\sigma_i^2 = \frac{S_e + S_p}{nk}. \tag{17}$$

Substituting the foregoing values of σ_s^2 and σ_i^2 in formula (9) gives, upon simplification,

$$r = \frac{S_p - \dfrac{S_e}{k-1}}{S_p + S_e}. \tag{18}$$

But

$$S_p = (n-1)M_p$$

and

$$S_e = (n-1)(k-1)M_e .$$

Substituting these values in formula (18) and simplifying gives

$$r = \frac{M_p - M_e}{M_p + (k-1)M_e} = \frac{M_{\bar{x}} - M}{M_{\bar{x}} + (k-1)M}.$$

In terms of these derivations, the average intercorrelation appears identical with the intraclass correlation formula. This will be true, however, only if the "between-raters" variance is not included as part of the error variance in either formula, or if it *is* included in both. As was previously pointed out, some situations require inclusion of the "between-raters" variance and others do not. The user of the intraclass formula finds it convenient to choose either procedure. The user of the average intercorrelation formula may also exercise an option in this matter. He may include the "between-raters" variance by calculating the average "within-raters" variance (σ_i^2) about the general mean for all raters. Or, he may exclude it by calculating that variance about the mean for each rater separately, as was indicated in the foregoing derivation.

It is also worth noting that the formulas give identical coefficients in spite of the fact that the average intercorrelation formula uses sample statistics whereas the intraclass formula uses estimates of population parameters. Since the reliability coefficient is basically a ratio of true and observed variances, application of the sample correction to both variances of the ratio does not change its value.

Consider finally the relation between the intraclass correlation formula and generalized formula for the reliability of averages. The latter, as given by Horst, is

$$r_{av} = 1 - \frac{\dfrac{\sum \dfrac{\sigma_i^2}{n_i - 1}}{N}}{\sigma_M^2}. \tag{19}$$

In the case where a complete table of ratings is available so that $n_1 = n_2 = n_3 = \cdots\cdots = n_i = k$, the formula may be written (substituting n for Horst's N)

$$r_{av} = 1 - \frac{\sum \sigma_i^2}{n(k-1)\sigma_M^2}. \tag{20}$$

The formula relating the reliability of an average of k equally weighted scores to the reliability of the scores themselves is

$$r_{av} = \frac{kr}{1 + (k-1)r}. \tag{21}$$

If the right hand members of equations (20) and (21) are equated and the equation solved for r, we obtain

$$r = \frac{\sigma_M^2 - \dfrac{\sum \sigma_i^2}{n(k-1)}}{\sigma_M^2 + \dfrac{\sum \sigma_i^2}{n}}. \tag{22}$$

Now since

$$\sigma_M^2 = \frac{\sum\left(\dfrac{P}{k}\right)^2}{n} - \left(\frac{T}{nk}\right)^2$$

or

$$nk^2 \sigma_M^2 = \sum P^2 - \frac{T^2}{n},$$

and since, as was shown previously,

$$k\,S_p = \sum P^2 - \frac{T^2}{n},$$

then

$$n\,k^2 \sigma_M^2 = k\,S_p$$

or

$$\sigma_M^2 = \frac{1}{n\,k} S_p.$$

Since, further,

$$\sum \sigma_i^2 = \frac{\sum\sum x^2}{k} - \sum\left(\frac{P}{k}\right)^2$$

or

$$k \sum \sigma_i^2 = \sum\sum x^2 - \sum \frac{P^2}{k},$$

and since

$$S_0 + S_e = S_t - S_p = \sum \sum x^2 - \frac{\sum P^2}{k},$$

then

$$k \sum \sigma_i^2 = S_0 + S_e$$

or

$$\sum \sigma_i^2 = \frac{S_0 + S_e}{k}.$$

Substituting the above values for σ_M^2 and $\sum \sigma_i^2$ in formula (22), we find

$$r = \frac{S_p - \dfrac{(S_0 + S_e)}{k-1}}{S_p + (S_0 + S_e)}. \tag{23}$$

Comparing this formula with formula (18) we observe that the sole difference is that the "error term" of formula (23) includes the "between-raters" sum of squares, whereas the formula (18) does not.

Hence, whenever there is a difference between the means of the ratings from various raters, the generalized reliability formula will give a lower value for the reliability of the ratings than is given by the intraclass formula as here calculated. The circumstances of each problem determine whether the "between-raters" variance should be included as part of the error term. As pointed out previously, the user of the intraclass formula may easily include or exclude "between-raters" variance as part of the error term. The user of the generalized reliability formula can not conveniently exclude "between-raters" variance even where it appears desirable to do so. In particular, the generalized reliability formula should not be used in estimating reliability of averages to which each rater has contributed by rating each product. Attention was called earlier to the fact that "between-raters" variance does not belong in the error term in this situation.

In this connection it should be mentioned that "between-raters" variance is always removed in the process of calculating the product-moment formula. Hence, as Gulliksen has pointed out in a private communication, the product-moment formula should not be used in cases where the "between-raters" variance is properly part of the error.

Both the intraclass formula and the generalized reliability formula are applicable to situations where the table of ratings is incomplete, that is, where k varies from product to product. But here again

a computational difference appears which results in different reliability coefficients. With the intraclass formula the mean square for error based on two component analysis is

$$Me = \frac{1}{N-n}\left[\Sigma\Sigma x^2 - \frac{\Sigma P^2}{k}\right], \tag{24}$$

so that each score contributes equally to the estimate of error variance. With the generalized reliability formula, however, the error variance for each product is figured separately and the separate error variances are then averaged, as indicated by the numerator of the fraction in formula (19). A "within-product" variance based on many scores is given the same weight as one based on few scores. Thus, a score in a small group is weighted more heavily than one in a large group.

A similar difference is observed in calculation of the variance of the means for products. In analysis of variance each product mean is weighted in proportion to the number of scores on which it is based. But in the generalized reliability formula each product mean is given equal weight. Since unweighted averages are seldom the same as weighted averages from the same data, the reliability estimates obtained when in the two formulas are applied to incomplete tables of scores are usually different.

Table 5 presents four sample problems, the data of which illustrate various combinations of equal or unequal means and variances. Below the data are given the reliability coefficients obtained by ap-

TABLE 5
Rater Reliability by Various Formulas for Problems Illustrating Various Conditions

Problem	1		2		3		4	
Condition								
Means	Equal		Unequal		Equal		Unequal	
Variances	Equal		Equal		Unequal		Unequal	
Scores (Test 1 and Test 2)								
Pupil 1	3	1	3	3	6.25	2	3	2
Pupil 2	1	3	1	5	4.25	6	1	6
Pupil 3	5	4	5	6	8.25	8	5	8
Pupil 4	4	5	4	7	7.25	10	4	10
Reliability								
Product-Moment	.4286		.4286		.4286		.4286	
Intraclass	.4286		.4286		.3429		.3429	
Average Intercorrelation	.4286		.4286		.3429		.3429	
Generalized Reliability	.4286		—.0196		.3429		—.0944	

plying each of the four formulas which have just been discussed. The values obtained confirm the analytical findings.

In Problem 1, with equal means and variances, all four formulas yield identical coefficients of reliability. In Problem 2, where only the means of the two sets of scores differ, the generalized reliability formula gives a much lower coefficient than the other three. This is due to the previously noted fact that systematic differences between the sets of scores are included as part of the error term in the generalized reliability formula, whereas these differences are removed in the other three formulas.

In Problem 3, where only the variances of the two sets of scores differ, all three of the special formulas yield coefficients which disagree with that from the product-moment formula, but which agree perfectly with each other. This is due to the fact that the product-moment formula uses a geometric mean of the variances of the scores from each observer, whereas all of the other formulas are based upon arithmetic means of these variances. In Problem 4, both the means and the variances of the two sets of scores differ, with the result that the intraclass coefficient and the average intercorrelation coefficient are somewhat lower, and the generalized reliability coefficient is very much lower, than the product-moment coefficient.

If the generalized reliability formula is applied to the data of Table 2 it yields a reliability coefficient of —.0582 compared with .1648 given by the intraclass formula. Here the discrepancy occurs because the generalized reliability formula uses a simple mean of the error variances within each person, and a simple variance of the person means, whereas the analysis of variance formula uses a weighted mean and a weighted variance.

The formula used by Clark was

$$r = \frac{(a\,\sigma_{av}^2/\sigma_i^2) - 1}{a - 1}. \tag{25}$$

The σ_{av}^2 is the variance of the averages for each product, and the σ_i^2 is, in this case, the total variance of all ratings. When this formula is applied to the data of Table 2 it yields a reliability coefficient of .2060 instead of the .1648 given by the intraclass formula. Here the discrepancy is partly due to the use in Clark's formula of a simple variance of the pupil means, and partly due to the estimation of the variance of the population of ratings from three sets of related ratings, which are treated as if they constituted a single random sample of the population of all ratings. Fisher (3, 225) has called attention to the error

introduced by the second of these procedures. These two differences in procedure no doubt cause much more apparent error in the small illustrative sample presented in Table 2 than they did in the more extensive data upon which Clark based his report. However the possibility of such errors when formula (25) is applied to incomplete sets of ratings should be recognized.

Conclusions

In view of the foregoing findings, there are three reasons for preferring the intraclass formula to the average intercorrelation or generalized reliability formulas. First, the intraclass formula permits the investigator to choose whether to include or exclude "between-raters" variance as part of the error variance, in terms of the circumstances of the particular problem. Second, a convenient means for estimating the precision of the reliability coefficients is available to the user of the intraclass formula. Third, the intraclass formula uses the familiar statistics and routine computational procedures of analysis of variance.

In the case of incomplete sets of ratings, the intraclass formula also has advantages over the generalized reliability formula and the variant of the average intercorrelation formula which Clark used in this situation. First, in determining the variances used in the intraclass formula, each observation is weighted equally, whereas in both the other formulas it is groups of unequal numbers of observations that are weighted equally. Second, Clark's formula involves a biased estimate of the population variance. Third, the advantages of the analysis of variance approach in estimating the precision of the reliability coefficient, and in computation, also apply in the case of unequal sets of ratings.

REFERENCES

1. Clark, Edward L. Reliability of college grades. Unpublished paper presented to American Psychological Association in September, 1950.
2. Cureton, Edward E. Reliability of a set of averages of unequal numbers of scores. Unpublished paper read at the March, 1951, meeting of the Southern Society for Philosophy and Psychology.
3. Fisher, R. A. Statistical methods for research workers. (10th Ed.) Edinburgh: Oliver and Boyd, 1946.
4. Gulliksen, Harold. Theory of mental tests. New York: John Wiley and Sons, 1950.
5. Horst, Paul. A generalized expression for the reliability of measures. *Psychometrika*, 1949, 14, 21-31.
6. Jackson, R. W. B., and Ferguson, George A. Studies on the reliability of tests. Toronto: Univ. of Toronto Press, 1941.

7. Lindquist, E. F. Statistical analysis in educational research. Boston: Houghton Mifflin Company, 1940.
8. Peters, Charles C., and Van Voorhis, Walter R. Statistical procedures and their mathematical bases. New York: McGraw Hill Book Company, 1940.
9. Snedecor, George W. Statistical methods. (4th Ed.) Ames, Iowa: Iowa State Coll. Press, 1946.

Manuscript received 11/30/50
Revised manuscript received 3/8/51

PSYCHOMETRIKA—VOL. 16, NO. 4
DECEMBER, 1951

A SQUARE ROOT METHOD OF SELECTING A MINIMUM SET OF VARIABLES IN MULTIPLE REGRESSION: II. A WORKED EXAMPLE

A. LUBIN
INSTITUTE OF PSYCHIATRY, MAUDSLEY HOSPITAL
AND
A. SUMMERFIELD
UNIVERSITY COLLEGE, LONDON

The *square root* method of selection has been explained in a previous article. In the present article a worked example is given which illustrates the compactness of the procedure. The square root method is compared with the Wherry-Doolittle method.

I. *Introduction*

The *square root* method of selection has been explained in a previous article (5). In the present article the technique is applied to the data of Paterson, Elliott, et al (4). Their data for a sample of 100 have also been employed by Garrett (2) to illustrate the Wherry-Doolittle procedure of test selection. We have applied the square root method to all ten tests to enable a comparison with the Wherry-Doolittle technique (see Section III) but to simplify the illustrative example we have only used those five tests with the highest beta weights (tests numbered 3, 4, 5, 7, and 9 by Garrett, and renumbered 1 to 5, respectively, by us). The square root method of selection is outlined step by step.

It is helpful to remember in what follows that the triangular, square root matrix, T, has the properties of a factor matrix, as was pointed out in (5). That is, (1) the sum of squares of any row of T is equal to unity; (2) the sum of the cross-products of any two rows of T gives, identically, the correlation between the tests corresponding to the two rows. Thus r_{34} in Table 1 equals:

$$r_{31}\, r_{41} + r_{3(2.1)}\, r_{4(2.1)} + r_{3(3.1,2)}\, r_{4(3.1,2)}$$
$$= .42(.56) + .136(.177) + .897(.224) = .46\,. \qquad (20)^*$$

The correlation matrix of predictors and criterion, R, is given in Table 1, there being 5 independent variables.

*The equations are numbered consecutively from our previous article (5).

TABLE 1

The Correlation Matrix, R, of the 5 Independent Variables and the Criterion ($N=100$)

Name of Test	Garrett's Numbers	No.	1	2	3	4	5	c
Minnesota Spatial Relations	3	1	1.00	.63	.42	.56	.55	.53
Paper Form Boards	4	2	.63	1.00	.37	.49	.61	.52
Stenquist Picture I	5	3	.42	.37	1.00	.46	.23	.24
Minnesota Assembly	7	4	.56	.49	.46	1.00	.41	.55
Interest Blank	9	5	.55	.61	.23	.41	1.00	.55
Quality Criterion	C	c	.53	.52	.24	.55	.55	1.00
Column Sums			3.69	3.62	2.72	3.47	3.35	3.39

TABLE 2

The Square Root Matrix, T

Name of Test	No.	Factors t_1 r_{i1}	t_2 $r_{i(2.1)}$	t_3 $r_{i(3.1,2)}$	t_4 $r_{i(4.1,2,3)}$	t_5 $r_{i(5.1,2,3,4)}$
Minnesota Spatial Relations	1	1.000	0	0	0	0
Paper Form Boards	2	0.630	0.777 √	0	0	0
Stenquist Picture I	3	0.420	0.136	0.897 √	0	0
Minnesota Assembly	4	0.560	0.177	0.224	0.778	0
Interest Blank	5	0.550	0.339	–0.052	0.069	0.758 √
Quality Criterion	c	0.530	0.240	–0.017	0.276	0.207
Column Sums	Σ	3.690 √	1.669 √	1.052 √	1.123 √	0.965 √
Cumulated Multiple R^2		0.281	0.339	0.339	0.415	0.458
Beta Weights		0.149	0.129	–0.086	0.331	0.273

II. *The Procedure*

A. *The complete square root matrix,* T, (see Table 2) *is first derived from the correlation matrix.*

Columns may be taken from R in any order, but most conveniently in their original order and so the first column of T, t_1, is identical with the first column of R. As t_1 accounts for all the variance of test 1, the first test has zero correlations with the remaining factors, t_2 to t_5.

t_2 is to account for the residual variance of test 2, so the correlation test 2 with t_2 *is*

$$r_{2(2.1)} = \sqrt{1-r^2_{12}} = \sqrt{1-(.63)^2} = .777\,. \tag{21}$$

The saturations of the other tests in t_2 are their semi-partial

correlations with test 2 adjusted against test 1, i.e.,

$$r_{i(2.1)} = \frac{r_{2i} - (r_{12})(r_{1i})}{r_{2(2.1)}}; \tag{22}$$

e.g., for test 5,

$$r_{5(2.1)} = \frac{.61 - .63(.55)}{.777} = .339,$$

and the criterion,

$$r_{c(2.1)} = \frac{.52 - .63(.53)}{.777} = .240.$$

Check: The column, t_2, may be checked as follows: Its sum should equal the result of applying equation (22), above, to the sums of the first two columns of R, instead of to the individual coefficients,

$$\sum(t_2) = \frac{\sum_i (r_{2i}) - (r_{12}) \cdot \sum_i (r_{1i})}{r_{2(2.1)}}; \tag{23}$$

i.e.,

$$1.669 = \frac{3.620 - .63(3.690)}{.777}.$$

The second column, t_2, of T has been made to account for the residual variance of test 2, which therefore has zero saturations in the remaining factors, t_3 to t_5.

t_3 is now derived in a similar way. The saturation of test 3 in t_3 is

$$\begin{aligned} r_{3(3.1,2)} &= \sqrt{1 - r^2_{13} - r^2_{3(2.1)}} = \sqrt{1 - (.42)^2 - (.136)^2} \\ &= .897. \end{aligned} \tag{24}$$

The saturations of the other tests in t_3 are their semi-partial correlations with test 3 adjusted against both test 1 and the adjusted test 2, i.e.,

$$r_{i(3.1,2)} = \frac{r_{3i} - (r_{13})(r_{1i}) - [r_{3(2.1)}][r_{i(2.1)}]}{r_{3(3.1,2)}}; \tag{25}$$

e.g., for test 5,

$$r_{5(3.1,2)} = \frac{.23 - .42(.55) - .136(.339)}{.897} = -.017.$$

Check: Carry out operation (25) on the sums, i.e.,

$$1.052 = \frac{2.720 - .42(3.690) - .136(1.669)}{.897}.$$

The process is similar for the remaining factors. As a final example taken from near the end of the calculation consider t_5. It has to account for the residual variance of test 5 after the removal of its saturations in factors t_1 to t_4. Its diagonal element is therefore:

$$[1 - (.55)^2 - (.339)^2 - (-.052)^2 - (.069)^2]^{1/2} = .758. \quad (26)$$

The saturation of the criterion in t_5 is its semi-partial correlation with test 5 adjusted against tests one to four:

$$r_{c(5.1,2,3,4)} = \frac{.55 - .55(.53) - .339(.240) - (-.052)(-.017) - (.069)(.276)}{.758} = .207. \quad (27)$$

The multiple R^2 of the criterion variable against the other 5 variables is the sum of squares of its saturations in factors t_1 to t_5, or its "communality" in these factors. The squares are cumulated in the bottom row of Table 2, their sum being finally 0.458, which is therefore $R^2_{c.1,2,\cdots,5}$.

$R^2_{c.1,2,\cdots,5}$ is tested for significance of difference from zero, (cf. 5, p. 279). The test is set out in Table 3 and the difference is found to be significant. Had it not been found significant the analysis would have ceased at this point.

TABLE 3
Test of the null hypothesis, $R^2_{c.1,2,3,4,5} = 0$

Source of Variance	d.f.	Sums of Squares	Mean Squares	F
Regression, i.e., $R^2_{c.1,2,3,4,5}$	5	0.458	0.0916	15.8*
Residual variance	94	0.542	0.0058	
Total	99	1.000		

*Significant; null hypothesis rejected.

B. *The calculation of the beta weights or multiple regression coefficients.*

There are five weights β_1 to β_5 and five equations are necessary

TABLE 4

Work Sheet: Selecting Variables by the Square Root Method*

Column No.: 1a	2a	3a	4a	5a	6a	7a	8a	9a
Name of Test	No.	r_{i4}	$1 - r^2_{i4}$	$r_{i(i.4)}$	r_{i5}	$r_{i(5.4)}$	(4a) $- r^2_{i(5.4)}$	$r_{i(i.4,5)}$
Minnesota Spatial Relations	1	0.560	0.686	0.828 √	0.550	0.351	0.563	0.750 √
Paper Form Boards	2	0.490	0.760	0.872 √	0.610	0.449	0.558	0.748 √
Stenquist Picture I	3	0.460	0.788	0.888 √	0.230	0.045	0.786	0.887 √
Minnesota Assembly	4	1.000	0	0 √	0.410	0	0	0 √
Interest Blank	5	0.410	0.832	0.912 √	1.000	0.912	0	0 √
Quality Criterion	c	0.550	0.698	0.835 √	0.550	0.356	0.571	0.756 √
Column Sums	Σ	3.470			3.350	2.113		
Column No.: 1b	**2b**	**3b**	**4b**	**5b**	**6b**	**7b**	**8b**	**9b**
Name of Test	**No.**	r_{ic}	$r_{ic} - r_{i4}r_{4c}$	$r_{c(i.4)}$			(4b) $- r_{i(5.4)}r_{c(5.4)}$	$r_{c(i.4,5)}$
Minnesota Spatial Relations	1	0.530	0.222	0.268			0.097	0.129
Paper Form Boards	2	0.520	0.250	0.287			0.091	0.121
Stenquist Picture I	3	0.240	-0.013	-0.015			-0.029	-0.033
Minnesota Assembly	4	0.550	0	0			0	0
Interest Blank	5	0.550	0.324	0.355			0	0
Quality Criterion	c	1.000	0.698	0.836			0.571	0.755
Column Sums	Σ	3.390	1.481				0.730	
Check Sums			1.482 √	1.481 √			0.731 √	0.729 √
Cumulated Multiple R^2		0.302		0.429				0.445

*To save space 3 decimal places have been used throughout this example. It would ordinarily be necessary to use a larger number to ensure accuracy.

TABLE 5

Tests of the null hypotheses: (A), $R^2_{c.4} = R^2_{c.4,5}$.
(B), $R^2_{c.4,5} = R^2_{c.1,4,5}$.
(C), $R^2_{c.4,5} = R^2_{c.1,2,3,4,5}$.

Test	Source of Variance	d.f.	Sums of Squares	Mean Squares	F
A	$R^2_{c.4}$ $[= r^2_{c4}]$	1	0.302		
	$R^2_{c.4,5} - R^2_{c.4}$ $[= r^2_{c(5.4)}]$	1	0.127	0.1270	21.5*
	Residual variance, $1 - R^2_{c.4,5}$	97	0.571	0.0059	
	Total	99	1.000		
B	$R^2_{c.4,5}$	2	0.429		
	$R^2_{c.1,4,5} - R^2_{c.4,5}$ $[= r^2_{c(1.4,5)}]$	1	0.017	0.0170	2.9†
	Residual variance, $1 - R^2_{c.1,4,5}$	96	0.554	0.0058	
	Total	99	1.000		
C	$R^2_{c.4,5}$	2	0.429		
	$R^2_{c.1,2,3,4,5} - R^2_{c.4,5}$	3	0.029	0.0097	1.7†
	Residual variance, $1 - R^2_{c.1,\cdots,5}$	94	0.542	0.0058	
	Total	99	1.000		

*Significant at 5% level; null hypothesis rejected.
†Non-significant at 5% level; null hypothesis accepted.

(See 5, pp. 276-278). From Table 2, using equation (9),

$$\beta_5 \, r_{5(5.1,2,3,4)} = r_{c(5.1,2,3,4)} ; \tag{28}$$

therefore, $\beta_5 = .207/.758 = .273$.

From equation (12) similarly

$$\beta_4 \, r_{4(4.1,2,3)} + \beta_5 \, r_{5(4.1,2,3)} = r_{c(4.1,2,3)} ; \tag{29}$$

therefore

$$\beta_4 = \frac{.276 - .273(.069)}{.778} = .331 .$$

If we continually use the "back solution" in this fashion, working from β_n to β_1, there is always only one unknown in each equation. For β_1 our solution is:

$$\sum_i \beta_i \, r_{i1} = r_{c1} , \tag{30}$$

or

$$\beta_1 = \frac{.53 - .129(.63) - (-.085)(.42) - .331(.56) - .273(.55)}{1.000}$$
$$= .149\,.$$

The check recommended (5, p. 277),

$$\sum_i \beta_i\, r_{ci} = R^2_{c(1,2,\cdots,n)}\,,$$

is now applied. The row of beta weights in Table 2 is multiplied by the row of criterion correlations in Table 1, i.e., we form the sum of cross products of criterion correlations with the corresponding beta weights and the sum should equal the multiple R^2, as it does.

$$.149(.53) + .129(.52) + (-.086)(.24) + .331(.55) + .273(.55) = .458\,.$$

If a desk calculating machine is used, it is possible to solve for each beta weight without setting down intermediate results.

C. *Selection of the effective predictors.*

Step 1: The variable having the highest validity is first selected. In the example, both tests 4 and 5 have correlations of 0.55 with the criterion. The decision between them is arbitrary* and test 4 has been chosen because this test was taken first by Garrett. Its column in R (Table 1) becomes t_1 (Column 3a, Table 4).

Step 2: The residual variance of the remaining tests are calculated each like $r^2_{2(2.1)}$ in A. above (Col. 4a, Table 4). The square roots of these quantities are entered in Col. 5a, Table 4. They are the denominators of the expression for the calculation of the semi-partial correlations of the criterion with tests 1, 2, 3, 5, and c separately, adjusted against test 4.

Check: Each row has to be checked separately; for each row the sum of squares of the elements in Cols. 3a and 5a should equal unity.

Step 3: The correlations of the criterion with tests 1, 2, 3, 4, 5 and c are entered in Col. 3b. Table 4. The numerators of the expressions for the semi-partial correlations of c with all tests are entered

*But if there is the possibility of improving the reliability of the tests, then, as Guilford (3, p. 423) points out, it is better to include the less reliable of two equally valid predictors, as a gain in validity is to be expected from an improvement in reliability. Burt (personal communication) has pointed out that, in general, it is better to take the variable with the lower average intercorrelation.

in Col. 4b. I.e., the quantities $r_{ic} - r_{c4}\, r_{i4}$; e.g., for test 3: [.24 — .55(.46)] = —.013 . The entry for the selected test, 4, will be zero.
Check: The sum of Col. 3b minus the product of r_{c4} (= .55) with the sum of Col. 3a should equal the sum of Col. 4b; i.e., carry out Step 3 upon the sums of the respective columns.

Thus:

$$[3.390 - .55(3.470)] = 1.482\,. \tag{31}$$

Step 4: The semi-partial correlations are computed by dividing each element of Col. 4b by the corresponding element of Col. 5a, the result being entered in Col. 5b.
Check: The sum of products of corresponding elements in Cols. 5a and 5 b should equal the sum of Col. 4b.

Step 5: The elements of Col. 5b are examined. The semi-partial correlation of c with test 5 is the largest. Had the largest semi-partial correlation with c been negative it would still have been selected; i.e., the largest absolute value, irrespective of sign, is chosen. The square of $r_{c(5.4)}$ is the difference between the multiple $R^2_{c.4,5}$ and $R^2_{c.4}$. This gain is tested (in Table 5A) and found to be significant. Test 5 is therefore extracted, and t_2 made to account for its residual variance.

The saturation of test 5 in t_2, $r_{5(5.4)}$, has already been calculated and appears as the element for test 5 in Col. 5a. It remains to calculate the semi-partial correlations of the remaining tests with the adjusted test 5, i.e., their saturations in t_2.

Step 6: The correlations, r_{5i}, of all tests are entered in Col. 6a. (As $r_{5(5.4)}$ is the denominator of the semi-partial correlations $r_{i(5.4)}$, use of its reciprocal simplifies subsequent calculations.) The semi-partial correlations $r_{i(5.4)}$ are calculated and entered in Col. 7a.

$$r_{i(5.4)} = \frac{r_{i5} - r_{45}\, r_{i4}}{r_{5(5.4)}}. \tag{32}$$

For test 3,

$$r_{3(5.4)} = \frac{.23 - .41(.46)}{.912} = .045.$$

Since it is unnecessary to record the numerator separately if a calculating machine is used, no column for it is provided in Table 4.
Check: Carry out operation (32) on the column totals:

$$\Sigma(\text{Col. 7a}) = \frac{[\Sigma(\text{Col. 6a}) - r_{45}\, \Sigma(\text{Col. 3a})]}{r_{5(5.4)}}$$

Tests 4 and 5 have now been selected. The cumulated multiple R^2 is entered in the bottom row of the lower half of Table 4. The five steps, 2 to 6, are the fundamental set of operations and are now repeated, with differences in detail, in the search for a further effective predictor.

Step 7 (cf. Step 2): The residual variances of all tests are entered in Col. 8a (cf. Col. 4a), from which they may be obtained by subtracting the squares of corresponding elements in Col. 7a. Col. 9a contains the square roots of these quantities.
Check: For each row, the sum of squares of elements in Cols. 3a, 7a, and 9a should equal unity.

Step 8 (cf. Step 3): The quantities

$$(r_{ic} - r_{c4}r_{i4} - r_{c(5.4)}r_{i(5.4)}) \tag{33}$$

are entered in Col. 8b. They are the numerators of the semi-partial correlations of c with all the tests. Col. 4b contains the quantities $(r_{ic} - r_{c4}\, r_{i4})$ and Col. 7a contains the semi-partial correlations $r_{i(5.4)}$. E.g., for test 3: $[-.013 - .571(.045)] = -.029$.

Check: Carry out operation (33) on the column sums:

$$\sum(\text{Col. 8b}) = \sum(\text{Col. 4b}) - r_{c(5.4)} \sum(\text{Col. 7a}).$$

Step 9 (cf. Step 4): The semi-partial correlations $r_{c(i.4,5)}$ are calculated by dividing the elements of Col. 8b by the corresponding elements of Col. 9a, and entered in Col. 9b.

Check: The sum of products of corresponding elements of Cols. 9a and 9b should equal the sum of Col. 8b.

Step 10 (cf. Step 5): The elements of Col. 9b are examined and $r_{c(1.4,5)} = .129$ is found to be the greatest. A test is therefore made of the difference between $R^2_{c.1,4,5}$ and $R^2_{c.4,5}$ (see Table 5b). The gain arising from the inclusion of test 1 is found to be non-significant. It follows that no other single test combined with tests 4 and 5 will add significantly to their multiple R^2 with the criterion, and the process of selection ceases at this point with $R^2_{c.4,5} = 0.429$.

(Had the contribution of test 1 proved significant, the second cycle of operations would have been completed with the further step (cf. Step 6) of calculating the semi-partial correlations, $r_{i(1.4,5)}$, i.e., t_3, and the search continued for a fourth effective independent variable.)

Step 11: It remains to test (*vide* the final paragraph of 5) whether some combination of the remaining tests might jointly add significantly to the multiple R^2, that is, to test whether $R^2_{c.4,5} = 0.429$ differs significantly from $R^2_{c.1,2,3,4,5} = 0.458$ (cf. A. above). The details of the test are given in Table 5c and the difference is found to be non-significant. The conclusions are therefore drawn that, under the conditions of the investigation, for $N = 100$, tests 4 and 5 are as effective as the complete battery in predictive power and that $R^2_{c.4,5} = 0.429$ is the effective multiple R^2 of the criterion with the battery.

If $R^2_{c.4,5}$ has been significantly lower than $R^2_{c.1,2,3,4,5}$, we would have continued adding variables until the multiple correlation based on the selected variates was not significantly different from the multiple correlation based on the whole battery.

It is perfectly possible for Steps 10 and 11 apparently to contradict each other, i.e., for Step 10 to indicate that no single variable when added to those already selected will produce a significant increment in the multiple correlation, and for Step 11 to indicate that the set of selected variables gives a multiple correlation that is significantly lower than the one obtained from the whole battery. This apparent contradiction, if it occurs, means that taking one variable at a time and adding its contribution to the battery is not sufficient. It would be necessary to take pairs (or higher combinations) of the non-selected, remaining variables and compare their joint contributions to the multiple correlation. An important question of methodology is involved here which has been discussed in the previous paper (5). (See discussion of Schützenberger's theorem on the "one-step locally best" solution.)

Step 12: *The calculation of the beta weights.* When a set of selected variables satisfying the criterion of Step 10 has been found, the beta weights should be calculated. The earlier paper (5) gives the underlying theory, and a fully worked example using the whole battery has been given above.

The square root matrix for the selected variables is needed. It

TABLE 6
The Triangular, Square Root Matrix of the Two Selected Variables

Name of Test	No.	r_{i4}	$r_{i(5.4)}$
Minnesota Assembly	4	1.000	0
Interest Blank	5	0.410	0.912
Quality Criterion	c	0.550	0.356

can be obtained from Columns 3a (r_{i4}) and 7a ($r_{i(5,4)}$) in Table 4.

Applying the same procedure as before, it is found that β_5 and β_4 both equal .390 .

As a check:

$$\beta_4 \, r_{c4} + \beta_5 \, r_{c5} = R^2_{c.4,5} = 0.429 \, ,$$

which is correct to the third decimal place.

The psychological interpretations are (1) A test of manual dexterity and a test of interest are the only variables necessary to account for that part of the criterion which is predicted by the whole battery, and (2) The two selected tests contribute equally to the prediction in this case, as they have equal correlations with the criterion, and therefore equal beta weights.

III. *Comparison of the Square Root Selection Method with the Wherry-Doolittle Method*

We have applied the square root method of test selection to the same set of ten variables used by Garrett (2) to illustrate the Wherdy-Doolittle method. The Wherry-Doolittle method led to the selection of four tests, tests 7, 9, 3, and 4 in that order giving a multiple $R^2_{C.7,9,3,4}$ of 0.452 . The square root method led to the selection of two tests only, tests 7 and 9, giving a multiple $R^2_{C.7,9}$ of 0.429. Test 3 was also found to have the highest residual validity after the extraction of tests 7 and 9, and the general pattern of the two methods is therefore seen to be similar. Test 3 was rejected by us, however, as not contributing significantly to the multiple R^2 of the criterion with tests 7 and 9. The multiple $R^2_{C.1,2,\cdots,10}$ of the criterion with the whole battery, was found to be 0.475. The F test showed that $R^2_{C.7,9}$ did not differ significantly from $R^2_{C.1,2,\cdots,10}$. In addition, as must obviously be the case, a further F test showed that Garrett's $R^2_{C.7,9,3,4}$ did not differ significantly from our $R^2_{C.7,9}$.

The reason for the discrepancy would appear to be that the Wherry-Doolittle method does not employ the 5% confidence level in deciding when to terminate the selection procedure. In effect, the Wherry-Doolittle method uses a non-parametric criterion. With this criterion, the null hypothesis, that an additional variable adds nothing to the multiple R^2, is accepted at a confidence level somewhere between 50% and our level of 5%. The criterion, based on the shrinkage of the multiple correlation, appears to be virtually identical with that proposed by Churchill Eisenhart (1), to which our comments extend.

Eisenhart's criterion is quoted by P. R. Rider (1, p. 126) as:

$$\frac{(1-R^2_{k+1})}{(1-R^2_k)} < 1 - \frac{1}{N-k-1}, \tag{34}$$

where k is the number of independent variables. Rewriting (34) as a variance ratio gives:

$$V = \frac{(1-R^2_{k+1})}{(N-k-2)} \cdot \frac{(N-k-1)}{(1-R^2_k)} < 1;$$
$$(\text{d.f.}: n_1 = N-k-2\,, n_2 = N-k-1) \tag{35}$$

i.e., the variance ratio of the residual variance with $k+1$ and k predictors. Equation (35) may be compared with the criterion that we have applied (cf. 5, pp. 278-280), which may be written:

$$F = \frac{(R^2_{k+1} - R^2_k)}{1} \cdot \frac{(N-k-2)}{(1-R^2_{k+1})},$$
$$(\text{d.f.}: n_1 = 1\,, n_2 = N-k-1) \tag{36}$$

and which we require to be significant (at the .05 level). Equation (36) we be rewritten as

$$F = \frac{[(1-R^2_k) - (1-R^2_{k+1})]}{1} \cdot \frac{(N-k-2)}{(1-R^2_{k+1})}. \tag{37}$$

Wherry (6, and as quoted by Garrett 2, cf. Table 65), effectively requires that:

$$\overline{K}^2_{m+1}/\overline{K}^2_m \leqslant 1\,, \tag{38}$$

where m is the number of independent variables and

$$\overline{K}^2_m = K^2_m \frac{(N-1)}{(N-m)} = (1-R^2_m)\,\frac{(N-1)}{(N-m)}. \tag{39}$$

The condition is, therefore, that

$$\frac{\overline{K}^2_{m+1}}{\overline{K}^2_m} = \frac{(1-R^2_{m+1})}{(N-m-1)} \cdot \frac{(N-m)}{(1-R^2_m)} \leqslant 1. \tag{40}$$

It will be seen that this expression is identical with Eisenhart's criterion, quoted above, excepting a difference in the degrees of freedom used. As the criterion variable absorbs one degree of freedom, in addition to those absorbed by the independent variables, Eisen-

hart's formulation would appear to be correct. For this reason, Eisenhart's rather than Wherry's version has been discussed.

Comparing (37) and (35), it will be seen that whereas we require the inclusion of a further independent variable to introduce a *significant* decrease in the residual variance term, to meet condition (35) merely requires that there shall be *some* reduction. It would therefore seem to follow that use of (34) as criterion for ceasing the selection of independent variables would necessarily lead to inflated estimates of the number of variables needed. The application of the F test to the difference between the results of the two methods in the present case reinforces this conclusion.

It is unlikely that Eisenhart's criterion was meant to be used as a test of significance.

REFERENCES

1 Eisenhart, C. Quoted from unpublished manuscript in Rider, P. R., Statistical Methods: New York: Wiley 1939 (p. 126).

2. Garrett, H. E. Statistics in psychology and education (3rd Ed.). New York: Longmans, 1947.

3. Guilford, J. P. Psychometric methods. New York: McGraw-Hill, 1936.

4. Paterson, D. G., Elliott, R. M., et al. Minnesota mechanical ability tests, Appendix 4. Minneapolis: University of Minnesota Press, 1930.

5. Summerfield, A., and Lubin, A. A square root method of selecting a minimum set of variables in multiple regression: I. The method. *Psychometrika*, 1951, 16, 271-284.

6. Wherry, R. J. A new formula for predicting the shrinkage of the coefficient of multiple correlation. *Ann. math. Statist.*, 1931, 2, 440-451.

Manuscript received 2/5/51

PSYCHOMETRIKA—VOL. 16, NO. 4
DECEMBER, 1951

BOOK REVIEWS

FLORENCE L. GOODENOUGH. *Mental Testing: Its History, Principles, and Applications.* New York: Rinehart and Co., 1950. Pp. xix+609.

Not too common are the college courses in psychology which are rich not only in the material offered but also in the ideas by which that material is organized and in the insights into the problems of the field. Dr. Florence Goodenough's course in mental testing, in which the reviewer was a student in 1937, was such a course. The content of that course is now recorded in book form with considerable additions.

The outstanding feature of this book is a history of intelligence testing from its beginnings to the present time. This history, which occupies the first six and parts of many later chapters, is both interesting and thorough. Another exceptionally interesting section is chapter 20, which gives rules for conducting an examination with young children, a subject on which Goodenough is an outstanding authority.

Most of Part II, "Principles and Methods," is lucid and valuable in understanding many aspects of the field of mental testing. There are many examples of that ability to see through the face value of statistical findings to the underlying meanings for which Goodenough is justly esteemed. Goodenough was one of the first to criticize the reliability coefficient; the dependence of reliabilities and other correlations on standard deviations is explained in a good footnote on page 164; the methodological intricacies in IQ constancy are well brought out; and so on.

Failure to bring the chapter on mental organization up to date is one of the book's disappointments. The chief protagonists in this chapter are Spearman and Thorndike, with the work of Kelley and Thurstone coming in for hardly more than mention. Fuller exposition of Thurstone's methods of factor analysis and of some of his results is found in Chapter 15, "Testing the Tests. I. General Principles and Fundamental Methods." In the last chapter of the book there is a brief contrast between McNemar's finding that one factor can account for most of the variance in the 1937 Stanford-Binet and Guilford's discovery of more than 20 factors in the Army Air Force testing program. The student would get a much clearer picture of the field of mental organization as it exists today if these materials had been brought together in a single chapter.

In the chapter on projective techniques, Freud and Jung are mentioned just once, Jung as having originated the free word association technique and Freud as having used it. Psychoanalysis is not mentioned. By contrast, four pages are devoted to Binet's *Experimental Study of Intelligence* with an additional page for pictures of his two daughters. Historically, however, the projective techniques did not arise in the Binet-testing tradition. The important people in the early history of the Rorschach test, the Thematic Apperception test, and play techniques were Freudian and Jungian analysts and psychotherapists with a background in psychoanalysis. A review of Freud's *Psychopathology of Everyday Life* would have been more germane to the history of the projective techniques and to the spirit of many of those who find the projective techniques useful than the review of Binet's study.

According to Goodenough, "The fundamental theory underlying all projective techniques is that every individual tends to project his own feelings and attitudes

upon the objects and people in the world by which he is surrounded." (p. 440) Both the usual scoring methods and the (more valid) clinical impressions based on projective techniques assume the operation of many mechanisms other than projection: displacement, condensation, denial, and so on. The student of this text gets no real idea of the complexity of the theory underlying interpretation of projective techniques.

The inventory type of personality test is the outcome of applying ability test techniques, the Binet tradition, to the field of personality. While Goodenough does not claim much for this type of test, it has a major fault which she does not mention: Extremely abnormal people may fail to be identified by questionnaires. Psychotics and psychopathic personalities may obtain normal ratings. Such extreme errors are much less likely with the use of projective techniques.

A disconcerting series of mistakes is found in the first paragraph on page 97. There the author states that Boring proposed an operational definition of intelligence as "what the tests test" at a symposium in Boston in 1921. The journal reference is to a symposium in the *Journal of Educational Psychology*. The definition is indeed Boring's, but it was made in the pages of the *New Republic* in 1923. Neither Boring nor any definition similar to his are to be found in the symposium in the *Journal of Educational Psychology*, nor does there seem to have been any physical meeting corresponding to this symposium. Strictly speaking, operationalism began some years later, with the publication of Bridgman's *Logic of Modern Physics* in 1927.

The book would be strengthened by the omission of Chapter 16, "Testing the Tests. II. The Divergence of Fact from Hypotheses," and Chapter 18, "Analysis of Variance."The methods presented in these chapters are fortunately not essential to the understanding of the main points in the rest of the book, and the exposition is not equal in clarity to that of the rest of the book. One error is the use of inverse probability, that is, assigning probabilities to hypotheses. By means of statistical analysis we assign probabilities to various possible outcomes of an experiment to test an hypothesis. When the actual outcome of the experiment is a probable one, our faith in the hypothesis is strengthened. When the actual outcome is an improbable one, we reject the hypothesis. Examples of assigning probabilities to hypotheses can be found on pages 249 and 270.

The discussion of the null hypothesis is even more confusing than is necessary in this difficult field. A statistical hypothesis is not just a negative statement, as Goodenough seems to imply. A statistical hypothesis ordinarily assigns some definite value to a parameter of a probability distribution. When that definite value happens to be zero, as it often is, we speak of the null hypothesis. Goodenough states, "It is possible to adduce evidence in support of a theory, whereas evidence against it cannot be regarded as proof that the theory is incorrect but only that its correctness has not been proved." (p. 232) Considering statistical hypotheses, Goodenough's assertion seems the opposite of the truth. A statistical hypothesis can be rejected, but it can never be accepted. If the experiment results in a highly improbable value for the test statistic, the hypothesis is rejected. If it results in a probable value, the hypothesis is not rejected, but it is not accepted either, for ordinarily there will be an infinite number of hypotheses which are also consistent with the experimental result.

There are errors in the presentation of formulas. The formula for χ^2 (p. 236) lacks a square sign. The formula for the correlation coefficient is incorrect on page 163 and again in the glossary. The formula for a T-score is given with indeter-

minate sign (p. 197). Many of the definitions of statistical terms found in the glossary are of doubtful value, for example, the definitions of t, F, degrees of freedom, and the phi coefficient.

As very little use is made of statistical results in evaluating tests, one wonders why so much space was expended introducing advanced statistical concepts. By contrast, Cronbach's competent book, *Essentials of Psychological Testing*, gives a minimum of statistical theory and a maximum of statistical results in relation to specific tests. Cronbach's tables of data about tests, however valuable to those able to make use of them, will undoubtedly throw many otherwise able clinicians into a state of "number shock." Goodenough's text, if revised to eliminate all but the simplest and most relevant statistical theory, will probably appeal more than Cronbach's to that type of student.

In summary, this text is outstanding in its discussion of the history and principles of intelligence tests. Projective tests are discussed unsympathetically. The statistical discussions range from illuminating to confusing, with the simple and more relevant topics discussed best. It is to be hoped that a revised edition of the book will soon be forthcoming, for it records the insights of one of the wisest and most fruitful contributors to the field of mental measurement in the period of that field's great growth.

Washington University *Jane Loevinger*

NIELS ARLEY and K. RANDER BUCH. *Introduction to the Theory of Probability and Statistics.* New York: John Wiley and sons, 1950. Pp. xi + 236. $4.00 (Translated from 1946 Danish edition)

J. NEYMAN. *First Course in Probability and Statistics.* New York: Henry Holt and Company, 1950. Pp ix + 350. $3.50.

The titles of these two books would suggest a greater similarity of content than actually exists, as can be best indicated by listing of chapter headings:

Arley and Buch:

1. The concept of probability (12 pages)
2. The foundations of the theory of probability (6 pages)
3. Elementary theorems [of probability] (7 pages)
4. Random variables and distribution functions (29 pages)
5. Mean value and dispersion (17 pages)
6. Mean value and dispersion of sums, products, and other functions (9 pages)
7. The normal distribution (21 pages)
8. Limit theorems (10 pages)
9. The relation of the theory of probability to experience and its practical applications (6 pages)
10. Application of the theory of probability to statistics (34 pages)
11. Application of the theory of probability to the theory of errors (29 pages)
12. Application of the theory of probability to the theory of adjustment [multivariate] (30 pages)

Neyman:

1. Introduction (14 pages)
2. Probability (81 pages)
3. Probabilistic problems in genetics (68 pages)

4. Random variables and frequency distributions (86 pages)
5. Elements of the theory of testing statistical hypotheses (95 pages)

Neyman keeps his discussion closer to probability and direct applications with greater emphasis on the logic of hypothesis testing, whereas Arley and Buch devote more to the topics ordinarily discussed in texts on mathematical statistics. Both books presuppose knowledge of integral calculus and both are written beyond the mathematical grasp of many students in psychology. Considered as possible reference books, Arley and Buch will be more useful than Neyman. The latter is superior on the general theory of statistical inference, and contains discussion of such topics as types of erroneous inferences, power function of a test, uniformly most powerful tests, and critical regions. It may be significant of something that in Neyman's *First Course* no mention is made of the mean and standard deviation.

Neither volume pretends to the more complete coverage of topics to be found in such texts as Hoel or Kenney or Mood or Cramér or Kendall. Although the Arley and Buch volume contains only two thirds as many pages as that of Neyman, they succeed in covering far more topics by relying more on the concise language of mathematics. In contrast, Neyman makes greater use of words.

Each volume may serve well its intended purpose, but it does not follow from this that either will be useful to more than a few students (or instructors) in psychology.

Stanford University *Quinn McNemar*

PSYCHOMETRIKA—VOL. 16, NO. 4
DECEMBER, 1951

PSYCHOMETRIC SOCIETY
STATEMENT OF RECEIPTS AND DISBURSEMENTS FOR FISCAL YEAR ENDED JUNE 30, 1951

RECEIPTS

Dues:

Membership Year	Members No.	Members Amt. Pd.	Student Members No.	Student Members Amt. Pd.		
1952	1	$ 5.00				
1951	328*	1640.00	47	$141.00		
1950	24	120.00	1	3.00		
1949	1	5.00				
	354	$1770.00	48	$144.00		
Total dues received						$1914.00
Other receipts, nature unknown						1.50
Total Receipts						$1915.50
DISBURSEMENTS						
Operating expenses						
Stationery and postage					$ 78.12	
Psychometric Corporation (90% of dues)					1722.60	
Total operating expenses						1800.72
Miscellaneous disbursements						
Checks returned to sender					10.00	
Postal note returned					3.00	
Total miscellaneous disbursements						13.00
Total Disbursements						$1813.72
BALANCE						
Excess of receipts over disbursements						$ 101.78
Balance as of June 30, 1950						994.33
Balance, June 30, 1951						$1096.11
Bank balance,† June 30, 1951						$1096.11

*Including applicants.
†Including checks totaling $13.00 entered in books and awaiting deposit.

PSYCHOMETRIC CORPORATION
STATEMENT OF RECEIPTS AND DISBURSEMENTS FOR FISCAL YEAR ENDED JUNE 30, 1951

RECEIPTS

Subscriptions:

Year	No.	Amt. Pd.
1950	29	$ 290.00
1951	305	3050.00
1952	5	50.00
1953	1	10.00
	340	$3400.00

PSYCHOMETRIC CORPORATION
STATEMENT OF RECEIPTS AND DISBURSEMENTS
FISCAL YEAR ENDED JUNE 30, 1951

RECEIPTS (Cont'd.)			
Total subscription payments received			$3400.00
Payments received for back volumes and issues			681.50
Total subscriptions and sales, at list price			4081.50
Less agency discounts			308.52
Net receipts from subscriptions and sales			3772.98
Receipts from Psychometric Society			1722.60
MISCELLANEOUS RECEIPTS			
Psychometric Monographs	$5.62		
Member's Dues	5.00		
Overpayment received	5.00		
Total miscellaneous receipts			15.62
Total Receipts			$5511.20
DISBURSEMENTS			
Operating expenses			
Dentan Printing Company—printing, mailing, etc.			$4159.85
Clerical and Secretarial Services			
Office of Editor		$600.00	
Office of Treasurer		162.66	
Total			$ 762.66
Stipend of Asst. Editor			375.00
Stationery and Postage			183.08
Treasurer's bond premium			25.00
Total operating expenses			$5505.59
Miscellaneous Disbursements			
Refund—Canceled subscriptions	$72.00		
Refund—Overpayments	11.25		
Adjustment on agency payments	1.50		
Transferred to Psychometric Society for dues	5.00		
Total miscellaneous disbursements			89.75
Total disbursements			$5595.34
BALANCE			
Balance as of June 30, 1950			$9229.53
Excess of expenditures over receipts			84.14
Balance, June 30, 1951			$9145.39
Bank Balance, June 30, 1951			$9145.39

INDEX OF VOLUME 16

first removes from each article the name of contributor and institution. The article is then sent to three or more persons who make independent judgments upon the suitability of the article submitted. This procedure seems to offer a possible mechanism for making judicious and fair selections.

Prospective authors are referred to the "Rules for Preparation of Manuscripts for Psychometrika," contained in the June, 1944 issue. Reprints of these "Rules" are available from the managing editor upon request. A manuscript which fails to comply with these requirements will be returned to the author for revision.

Under ordinary circumstances authors will not be charged for cuts, tables, formulas, or unusual type faces. Authors will receive 200 reprints without covers, free of charge.

Manuscripts for publication in *Psychometrika* should be sent to

DOROTHY C. ADKINS, Managing Editor, *Psychometrika*
Department of Psychology, The University of North Carolina
Chapel Hill, North Carolina

Material for review in *Psychometrika* should be sent to

CLYDE H. COOMBS, Review Editor, *Psychometrika*
Department of Psychology, The University of Michigan
Ann Arbor, Michigan

The officers of the Psychometric Society for the year September 1951 to September 1952 are as follows: *President*: John C. Flanagan, American Institute for Research, University of Pittsburgh, Pittsburgh, Pennsylvania; *Secretary*: Clyde H. Coombs, The University of Michigan, Ann Arbor, Michigan; *Treasurer*: Robert L. Thorndike, Teachers College, Columbia University, New York, New York.

The Council members, together with dates at which terms expire, are as follows: Hubert Brogden, 1954; Philip DuBois, 1954; Paul Horst, 1953; Herbert S. Conrad, 1952; Quinn McNemar, 1952.

The *Psychometric Monographs Committee* is composed of J. P. Guilford, *Chairman*; L. L. Thurstone, Harold Gulliksen, Paul Horst, and Frederic Kuder. Manuscripts and correspondence for this series should be addressed to

J. P. GUILFORD, Chairman, Psychometric Monographs Committee,
Box 1134, Beverly Hills, California

PRINTED BY THE DENTAN PRINTING CO., COLORADO SPRINGS, COLORADO